

Modellbasierte Lokalisation und Verfolgung für sichtsistemgestützte Regelungen

Zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften

von der Fakultät für Informatik
der Universität Karlsruhe (TH)

genehmigte

Dissertation

von

Thomas Müller

aus

Walldorf bei Heidelberg

Tag der mündlichen Prüfung:

8. Februar 2001

Erster Gutachter: Prof. Dr. H.-H. Nagel

Zweiter Gutachter: Prof. Dr. H. Prautzsch

Danksagung

An erster Stelle danke ich natürlich Herrn Prof. Dr. H.-H. Nagel für die außerordentlich engagierte Betreuung meiner Arbeit, für all seine Bemühungen sowie die guten Ratschläge. Von ihm habe ich nicht nur viel über wissenschaftliches Arbeiten gelernt.

Ganz besonderen Dank möchte ich Herrn Prof. Dr. H. Prautzsch aussprechen für die Übernahme des Korreferats trotz enger zeitlicher Rahmenbedingungen. Dank seiner freundlichen Unterstützung gelang es mir, die Abhandlung zügig fertigzustellen und bereits im Wintersemester 2000/2001 zu promovieren. Herzlichen Dank.

Für das angenehme kollegiale und freundschaftliche Verhältnis bedanke ich mich sehr herzlich bei meinen Kollegen Dr. Holger Leuck, Dr. Ralf Gerber, Ferit Keçeci, Alexander Bachem, Klaus Fleischer, Markus Middendorf und Dr. Clemens Ballarin sowie bei meinen ehemaligen Kollegen Dr. Michael Haag, Dr. Frank Heimes, Dr. Klaus Mück, Angela Gehrke und Dr. Karl Schäfer. Allen möchte ich meinen Dank aussprechen für die Unterstützung moralischer oder technischer Art sowie für jegliche wissenschaftliche Anregungen. Besonderer Dank geht dabei an Dr. Holger Leuck für seine freundliche Unterstützung sowie für die wissenschaftlichen und nicht wissenschaftlichen Diskussionen. Besonders danken möchte ich auch Dr. Michael Haag, der mir anfangs den Einstieg in die Forschungstätigkeit deutlich erleichterte.

Bedanken möchte ich mich bei Dr. Volker Gengenbach für seine Hilfe bei einigen scheinbar unüberwindlichen programm- oder gerätetechnischen Schwierigkeiten. Ferner danke ich Dr. Stefan Wachter für die Überlassung seines Programmsystems zur Personenverfolgung und für die Einweisungen dazu.

Besonderen Dank möchte ich Frau Ulla Dietrich aussprechen für all die freundlichen Worte und Taten, die Licht in den manchmal dunklen Alltag brachten.

Ganz besonders herzlich danke ich vor allem meinen Eltern, meiner Schwester, meinem Bruder und meinen Freunden, die mir halfen, die manchmal schwere Zeit zu überstehen, die mir Mut machten und mich mit Rat und Tat unterstützten.

Kurzfassung

Der Sehsinn des Menschen bildet ein mächtiges Hilfsmittel zur Erfassung, Beobachtung und Interpretation seiner Umgebung und damit eine wichtige Grundlage zur Interaktion mit der Umwelt. So ist es nicht verwunderlich, daß mit einer oder mehreren Kameras ausgestattete Sichtsysteme in vielen Einsatzbereichen ebenfalls eine – mit anderen Sensoren oft unerreichte – zentrale und damit auch besonders zukunftssträchtige Rolle spielen.

In der vorliegenden Arbeit wird anhand zweier bewußt sehr unterschiedlich gewählter Teilbereiche das Feld *modellbasierter Sichtsysteme* näher untersucht. Bei dem einen handelt es sich um die rechnergestützte Erkennung der Bewegungen von Personen. Der zweite Bereich beschäftigt sich mit einer sichtsystemgestützten Roboterregelung für eine komponentenweise *Demontage* des Motorraums von Altfahrzeugen, die im Gegensatz zu einer automatischen *Montage* wesentlich auf Sichtsysteme angewiesen ist. Die Verschiedenartigkeit der beiden Teilbereiche ermöglicht dabei, die einzelnen relevanten Aspekte modellbasierter Sichtsysteme durch Betrachtung des jeweils dafür günstigeren Bereichs besonders gut herauszuarbeiten, und bedeutet gleichzeitig eine breite Abdeckung des Feldes modellbasierter Sichtsysteme.

Ziel der vorliegenden Arbeit ist eine signifikante Verbesserung hinsichtlich der in Theorie und Praxis relevanten Aspekte, wie z.B. Automatisierbarkeit, Robustheit, Genauigkeit, Flexibilität, Ausführungsgeschwindigkeit, Zuverlässigkeit und Analysierbarkeit. Dies bedeutet zum einen die *Entwicklung neuer Verfahren* und zum anderen die Analyse und *Verbesserung zahlreicher Einflußfaktoren*. Letzteres zeigt sich beispielsweise insbesondere bei der Steigerung der Robustheit als unumgänglich. Beide Vorgehensweisen werden in der vorliegenden Arbeit behandelt und zu allen genannten Aspekten *signifikante Fortschritte* erzielt. Dazu zählen insbesondere die Erweiterung der Demontage um eine automatische, Realwelt-taugliche, mono- oder multiokulare *Objektlokalisierung* zur Initialisierung des zugrundeliegenden Verfolgungsprozesses, die nun *echtzeitfähige* und hinreichend *robuste* Roboterregelung, eine *bildfolgenunabhängige* Wahl der Parameter bei der Personenverfolgung sowie der Einsatz eines iterativ erweiterten Kalmanfilters unter Verwendung einer Levenberg-Marquardt–Optimierung bei der Personenverfolgung (anstatt eines erweiterten Kalmanfilters mit Gauß-Newton–Optimierung).

Inhaltsverzeichnis

1	Einleitung	1
1.1	Personenverfolgung	3
1.2	Sichtsystemgestützte Demontage	4
1.3	Ziel der vorliegenden Arbeit	9
2	Literaturüberblick	11
2.1	Erkennung und Lokalisation	11
2.1.1	Merkmalsbasierte Verfahren	13
2.1.2	Freiform <i>kurven</i> basierte Verfahren	18
2.1.3	Freiform <i>flächen</i> basierte Verfahren	19
2.1.4	Überführung von Punktmengen ineinander	19
2.1.5	Gebietsbasierte Verfahren, Schablonenanpassung	20
2.1.6	Indizierungsbasierte Verfahren	21
2.1.7	Verfahren mittels Hough-Transformation	23
2.1.8	Betrachtungswinkelbasierte Verfahren	24
2.1.9	Probabilistische Verfahren	26
2.1.10	Neuronale Netze verwendende Verfahren	28
2.1.11	Sonstige Verfahren	29
2.1.12	Zusammenfassung und Diskussion	30
2.2	Personenverfolgung	31
2.3	Verfolgung/Regelung	32
2.4	Bewertungsaspekte	34
3	Personenverfolgung	37
3.1	Genauigkeit und Robustheit	38

3.1.1	Definition von Genauigkeit und Robustheit	38
3.1.2	Chaotischer Schwingungseffekt	41
3.1.3	Einsatz eines iterativ erweiterten Kalmanfilters unter Verwendung einer Levenberg-Marquardt-Optimierung	46
3.1.4	Freiheitsgradwertebeschränkungen	52
3.1.5	Zusammenfassung und Ausblick	60
3.2	Automatisierbarkeit	62
3.2.1	Übergang von einer bildfolgenindividuellen zu einer bildfolgenunabhängigen Parametrisierung	64
3.3	Analysierbarkeit, Zuverlässigkeit, Ausfallsicherheit	90
4	Sichtsystemgestützte Demontage	92
4.1	Lokalisation zur Verbesserung der Automatisierung	93
4.1.1	Such- und Winkelbereiche zu den projizierten Modellkanten	96
4.1.2	Bewertungsfunktion für die einzelnen Modellagen	98
4.1.3	Verfeinerung der Suche	99
4.1.4	Zählung der geeigneten Kantenelemente in den Suchbereichen	102
4.1.5	Multiokulare Lokalisation	104
4.1.6	Berücksichtigung von a-priori Wissen zur Objektlage	105
4.1.7	Ergebnisse	105
4.2	Robustheit und Genauigkeit	114
4.2.1	Verwendung einer adaptiven Bildfilterung	115
4.2.2	Verbesserung verschiedener Einzelfaktoren	122
4.3	Zuverlässigkeit, Sicherheit, Ausfallsicherheit	124
4.4	Flexibilität	129
4.4.1	Flexible, universelle Einsetzbarkeit	129
4.4.2	Regelung eines Schraubers	130
4.4.3	Verfolgung weiterer Objekte	133
4.4.4	Ansteuerung frei wählbarer räumlicher Bahnkurven des Beobachtungsroboters	136
4.5	Geschwindigkeit	140
4.5.1	Steigerung der Echtzeitfähigkeit bei der Regelung	142
4.5.2	Verteilte Berechnung der Lokalisation	148
4.6	Analysierbarkeit	153

5	Gegenüberstellung der beiden Teilbereiche	154
6	Zusammenfassung und Ausblick	158
	Literaturverzeichnis	162
	Index	172

Abbildungsverzeichnis

1.1	Referenzversuche: Verfolgung bei den sechs Bildfolgen nach [Wachter 97]; Bildfolgennamen	5
1.2	Versuchsumgebung Roboterzelle	6
1.3	Kantenelementbild zu einer Aufnahme der Autobatterie	7
1.4	Monokulare Verfolgung der Autobatterie zum Halbbildzeitpunkt 3 sowie die im Anpassungsprozeß verwendeten Kantenelemente	8
1.5	Stereoverfolgung des Luftfilters	8
1.6	Regelung des Greifers an die korrekte Greifposition und Entnahme der Batterie aus dem Motorraum	9
3.1	Modellanpassungen zu einer Verfolgung zu <code>alinde</code> zu den Halbbildzeitpunkten 330 bis 333	42
3.2	Prädizierte Position der Person in x -Richtung und Geschwindigkeit in x -Richtung bei der Verfolgung zu <code>alinde</code>	43
3.3	Kleinste Eigenwerte und Konditionszahlen der in den Aktualisierungsschritten zu einer Verfolgung mittels <code>hof_191096</code> auftretenden approximierten Hessematrizen	49
3.4	Modellanpassungen des Referenzversuchs und bei Versuch 6 zum Halbbildzeitpunkt 209	51
3.5	Modellanpassungen bei <code>frankm</code> zu den Halbbildzeitpunkten 281 und 282	53
3.6	Modellanpassungen des Referenzversuches sowie bei Versuch 7 und 8 bei <code>frankm</code> zum Halbbildzeitpunkt 281	56
3.7	Prädizierter Winkel für die Beugung des linken Knies und die zugehörige Winkelgeschwindigkeit im Referenzversuch und in Versuch 8 zu <code>frankm</code>	57
3.8	Modellanpassungen bei einer Verfolgung zu <code>s1_191096</code> zu den Halbbildzeitpunkten 328 und 333. Die Bewegung des linken Armes wird aufgrund des Apertureffekts falsch eingeschätzt	62

3.9	Modellanpassungen zu frankm zu den Halbbildzeitpunkten 96 und 104. Das linke Modellbein wird dem rechten mit angepaßt	63
3.10	Modellanpassung bei einer Verfolgung zu hof_191096 zum Halbbildzeitpunkt 79. Es entsteht ein Verfolgungsproblem aufgrund des niedrigen Kontrasts zwischen den Beinen und dem Hintergrund	63
3.11	Modellanpassungen von Versuch 33 zu den Halbbildzeitpunkten 86 und 176	74
3.12	Modellanpassungen der Versuche 33 und 38 zum Halbbildzeitpunkt 161	76
3.13	Modellanpassungen der Versuche 33 und 38 zum Halbbildzeitpunkt 53 .	76
3.14	Modellanpassungen der Versuche 43 und 68 zum Halbbildzeitpunkt 462 zur Demonstration der Wirkung des Parameters Zylinderenden komplett	84
4.1	Suchbereich zu den projizierten Modellkanten bei der Objektlokalisierung	97
4.2	Veranschaulichung der bei den Iterationen der hierarchischen Suche zu bewertenden Modellagen, den verwendeten Granularitäten und die bei der Bestimmung der Such- und Winkelbereiche zu den einzelnen projizierten Modellkanten relevanten Größen sowie der Zusammenhang zwischen diesen Werten	100
4.3	Monokulare Lokalisation der Batterie, des Luftfilters, des Unterdruckbehälters sowie der mit Schrauben verbundenen Metallquader	108
4.4	Binokulare Lokalisation der Batterie und des Luftfilters	109
4.5	Binokulare Lokalisation des Unterdruckbehälters und des Schiffsteils . .	110
4.6	Ergebnisse der Lokalisationen des Schiffsteils in den hochauflösenden Bildern 1 und 2	111
4.7	Ergebnisse der Lokalisationen des Schiffsteils in den hochauflösenden Bildern 3 und 4	112
4.8	Ergebnisse der Lokalisationen des Schiffsteils in den hochauflösenden Bildern 5 und 6	113
4.9	Bild der Autobatterie und extrahierte Kantenelemente bei Verwendung der Maskenberechnung mittels Abtastung einer abgeleiteten Gaußfunktion und der Maskenberechnung mit Randausgleich nach [Otte 94] bei einer Filterung mit konstanten Masken	116
4.10	Adaptive Bildfilterung mit Eigenwerttransformation nach den Gleichungen (4.4) und (4.6)	117
4.11	Adaptive Bildfilterung mit Eigenwerttransformation nach den Gleichungen (4.7) und (4.8)	118

4.12	Adaptive Bildfilterung mit Eigenwerttransformation nach den Gleichungen (4.9) und (4.10) für $x_{\min} = 10$, $x_{\max} = 11$	119
4.13	Verfolgung der Batterie zum Halbbildzeitpunkt 201 bei der bisherigen Bildfilterung mit ortsinvarianten Filtermasken und der adaptiven Bildfilterung nach (4.9), (4.10) mit $x_{\min} = 10$, $x_{\max} = 11$ bei einer Schwelle für den Minimalabstand projizierter Modellkanten von 2,75 Pixel	122
4.14	Modell des Greiferwerkzeuges vor und nach einer Erweiterung	124
4.15	Verschaltung der Spezialrechner (Transputer) zur Geräteansteuerung .	126
4.16	Verbesserte Verschaltung der Spezialrechner (Transputer) zur Geräteansteuerung	127
4.17	Doppelklotz aus [Gengenbach 94]	130
4.18	Lageschätzung des Doppelklotzes bei der Regelung des Schraubers zu den Halbbildzeitpunkten 1, 810, 1180 und 1439	132
4.19	Lage des kugelförmigen Behälters im Motorraum des Versuchsfahrzeugs	134
4.20	Monokulare Verfolgung des kugelförmigen Behälters	134
4.21	Stereoverfolgung kugelförmigen Behälters und des Schiffsteils	135
4.22	Splinekurve mit Stütz- und Kontrollpunkten für die Fahrt des Beobachtungsroboters	137
4.23	Spezielle Splinekurve für die Kreisbewegung des Beobachtungsroboters mit Stütz- und Kontrollpunkten	140
4.24	Ergebnis der Anpassung des automatisch lokalisierten Schiffsteils bei einer Kreisbewegung der Kamera zum Beginn der Kreisfahrt, nach einem Drittel und nach zwei Dritteln der Fahrt	141
4.25	Datenfluß- und Ablaufdiagramm der prozeß(faden)orientierten monokularen Verfolgung nach [Tonko 97]	143
4.26	Realisierte Prozedurstruktur bei der monokularen Verfolgung	144
4.27	Zusammenspiel und Ablauf der drei Prozesse bei der Stereoverfolgung nach [Tonko 97]	146
4.28	Realisierte Prozedurstruktur bei der binokularen Verfolgung	147
4.29	Sämtliche relevante Geräte des Instituts	149
4.30	Abhängigkeit der Lokalisationszeit bzw. -rate von der Leistung der verwendeten Gerätekonfiguration bei einer Lokalisation des Schiffsteils . .	151
5.1	Positionsschätzungen $(x, y, z)^T$ bei Verfolgung der Batterie, des Luftfilters sowie des kugelförmigen Behälters während zwei Bewegungsperioden der Kamera	157

Tabellenverzeichnis

3.1	Abweichungen der Matrizen P_{uu} und P_{vv} von ihren Transponierten bei der Verfolgung zu <code>alinde</code>	44
3.2	Fehler in der Symmetrie der Kovarianzmatrizen \hat{P}_k^- zu einigen Halbbildzeitpunkten bei einer Verfolgung zu <code>hof_191096</code>	46
3.3	Versuche 1 bis 6 zum iterierten erweiterten Kalmanfilter mit Levenberg-Marquardt–Minimierung	51
3.4	Zusammenstellung aller wesentlichen Unterschiede der Ergebnisse der Versuche 1 bis 6 im Vergleich zu den Referenzversuchen	52
3.5	Beugewinkel des linken Knies und dessen Winkelgeschwindigkeit bei <code>frankm</code> zu den Zeitpunkten 280 bis 282, wenn die Beschränkung des Winkels ab Halbbild 280 ein- bzw. abgeschaltet ist	54
3.6	Versuche 7 und 8 zu <code>frankm</code> unter Verwendung der Extremwerte $-0,1576$ <i>rad</i> bzw. $-0,1503$ <i>rad</i> für den Beugewinkel des linken Knies	55
3.7	Versuche 9 bis 14: Wiederholung der Versuche 1 bis 6 ohne Geschwindigkeitsrücksetzung bei den Freiheitsgradwertebeschränkungen	60
3.8	Bisher eingesetzte Projektionsparameter bei den Bildfolgen	65
3.9	Bisher eingesetzte Verfolgungsparameter bei den Bildfolgen	65
3.10	Versuche 15 bis 20 zur Angleichung des Modellgradientenverfahrens bei den Bildfolgen	68
3.11	Versuche 21 bis 26 zur Angleichung des Modellgradientenverfahrens bei den Bildfolgen	70
3.12	Versuche 27 bis 30 zur Angleichung der Parameter <code>Kantenabstand</code> und <code>Flächenabstand</code>	71
3.13	Versuche 31 bis 36: Wiederholung der Versuche 21, 27, 28, 29, 5 und 30 unter Verwendung von <code>Kantenabstand = Flächenabstand = 2</code> zur Parameteroptimierung	73
3.14	Versuche 37 und 38: Wiederholung von Versuch 33 unter Verwendung breiterer Modellunterschenkel	75

3.15	Versuche 39 bis 45: Wiederholung der Versuche 31 bis 36 bei verändertem Parameter Meßrauschen Flächen	79
3.16	Versuche 46 bis 51: Wiederholung der Versuche 31, 40, 38, 34, 43 und 45 bei Meßrauschen Flächen = 14	80
3.17	Versuche 52 bis 57: Wiederholung der Versuche 31, 40, 38, 34, 43 und 45 bei Meßrauschen Flächen = 7	81
3.18	Versuche 58 bis 63: Wiederholung der Versuche 31, 40, 38, 34, 43 und 45 unter Verwendung von Kantenabstand = Flächenabstand = 1 .	84
3.19	Versuche 64 bis 69: Wiederholung der Versuche 31, 40, 38, 34, 43 und 45 bei geänderten Parameter Zylinderenden komplett	86
3.20	Versuche 70 bis 75: Wiederholung der Versuche 31, 65, 66, 34, 43 und 45 bei geänderten Parameter maximale Grauwertdifferenz für Flächenmessung	89
3.21	Angeglichene und optimierte Parameter bei den Bildfolgen	89
4.1	Lokalisationsparameter aller mono- und binokularen Lokalisationen . .	107
4.2	Verbesserung des Visdom -Systems durch die Bereinigungsaktionen, quantifiziert anhand einiger Meßgrößen	128
4.3	Schritte bei der Verbesserung der Echtzeitanforderungen an die monokulare Verfolgung, gemessene Zyklus- und Latenzzeiten sowie die bei jedem Schritt erzielte Verbesserung	144
4.4	Mittlere Zyklus- und Latenzzeiten sowie durchschnittliche Dauern der einzelnen Programmteile für alle monokularen und binokularen Verfolgungen und die Demontagevorgänge	148
4.5	Zur Verfügung stehende Rechner mit einer Kapazität über 150 MHz . .	150
4.6	Mono- und binokulare Lokalisationszeiten abhängig von der Rechnerkonfiguration	152

Kapitel 1

Einleitung

Zweifelloos bildet der Sehsinn des Menschen ein mächtiges Hilfsmittel zur Erfassung, Beobachtung und Interpretation seiner Umgebung und damit eine wichtige Grundlage zur Interaktion mit der Umwelt. Will man künstliche Systeme realisieren, die menschliche Fähigkeiten auf Interpretations- oder Interaktionsebene nachempfinden, gilt naturgemäß Entsprechendes. So ist leicht nachzuvollziehen, daß mit einer oder mehreren Kameras ausgestattete Sichtsysteme in vielen Einsatzbereichen eine – mit anderen Sensoren oft unerreichte – zentrale und damit auch besonders zukunftssträchtige Rolle spielen.

Aufgrund der Wichtigkeit solcher Systeme wird in der vorliegenden Arbeit das Feld *modellbasierter Sichtsysteme* näher untersucht. Die Nutzung von Objektmodellen – etwa in Form dreidimensionaler CAD- oder Volumenmodelle – erweist sich bei Sichtsystemen als besonders günstig, da dies in geeigneter und kompakter Form die zu beobachtenden Objekte zu beschreiben vermag und so Rückschlüsse auf die Verhältnisse in der *3D-Szene* erlaubt (im Gegensatz zu Verfahren, die im *2D-Bildbereich* arbeiten und charakteristischerweise lediglich Aussagen über die Verhältnisse im Kamerabild treffen können).

Die Untersuchungen innerhalb des Betätigungsfeldes erfolgen anhand zweier bewußt sehr unterschiedlich gewählter Teilbereiche. Dies ermöglicht es, die einzelnen relevanten Aspekte modellbasierter Sichtsysteme durch Betrachtung des jeweils dafür günstigeren Bereichs besonders gut herauszuarbeiten. Gleichzeitig wird durch die unterschiedlichen Zielsetzungen eine breite Abdeckung innerhalb des Feldes modellbasierter Sichtsysteme erreicht. Bei beiden Teilbereichen kann auf einem bereits bestehenden Programmsystem aufgebaut werden, das aktuelle Standards erfüllt. Dies erleichtert bei den Untersuchungen eine Anknüpfung an den Stand der Forschung.

Bei dem einen Teilbereich handelt es sich um die rechnergestützte Erkennung der Bewegungen von Personen. Hierbei muß die prinzipiell *unvorhersehbare* Bewegung *mehrerer Körperteile* in ständig wechselnden Situationen unter Berücksichtigung gegenseitiger partieller oder totaler Verdeckungen erkannt werden. Zentrales Problem ist die Weiter-

entwicklung zu einem hinreichend robusten Verfahren, das die durchgeführten Personenbewegungen trotz der Komplexität der zu lösenden Aufgabe zuverlässig erkennen kann. Hierbei ist es sinnvoll, Geschwindigkeitsaspekte demgegenüber zurückzustellen, solange kein hinreichender Grad an Robustheit erreicht worden ist.

Der zweite Bereich beschäftigt sich mit einer sichtsystemgestützten Roboterregelung für eine komponentenweise *Demontage* des Motorraums von Altfahrzeugen. Während die automatische *Montage* von Fahrzeugen in einer Fabrik meist ohne Sichtsensorik mittels Vorgabe von festen Roboterzielpositionen bewerkstelligt werden kann, kommt eine komponentenorientierte *Demontage* von Altfahrzeugen nicht ohne Kameras aus, da im allgemeinen viele unterschiedliche Fahrzeugmodelle und -typen (möglichst ohne rechnerinterne Repräsentation jedes einzelnen Typs), sowie Spezialkonstruktionen, Umbauten und Unfallwagen behandelt werden müssen. Das bedeutet, daß ein solches Demontagesystem in der Lage sein muß, Komponenten unerwarteter Lage, Größe oder Form zu handhaben.

Im Vergleich zur Personenverfolgung müssen bei diesem System nur wenige Parameter zur Verfolgung *weniger Objekte* aus dem Bildmaterial geschätzt werden (da *starre* Körper verfolgt werden und zur Regelung eines Werkzeugs bereits eine Verfolgung von nur zwei Objekten (Werkzeug und Werkstück) ausreicht). Die Objektbewegung wird zudem hauptsächlich vom Demontagesystem selbst bewirkt, was die Veränderungen im aufgenommenen Bildmaterial bis zu einem bestimmten Grad *vorhersehbar* macht. Die Komplexität der zu lösenden Aufgabe entsteht auf diesem Teilgebiet hauptsächlich durch die Verschiedenartigkeit der Objekte innerhalb komplexer Realwelt-Szenen und durch die benötigten Anforderungen an die Demontageaktionen. Zudem stehen besonders auch *Echtzeitaspekte* im Vordergrund. Da aber robuste Algorithmen zur Bewältigung der Aufgabenkomplexität im Allgemeinen zeitaufwendig sind, muß oft ein Kompromiß gefunden werden zwischen der Robustheit auf der einen Seite und der Ausführungsgeschwindigkeit auf der anderen Seite.

Während also bei der Personenverfolgung die Robustheit ungeachtet der dazu benötigten Rechenzeit zu steigern ist, muß bei der Demontage die Robustheit möglichst ohne Rechenzeitverluste bzw. die Geschwindigkeit möglichst ohne Robustheitsverluste verbessert werden. Dieser Punkt ist bei der Übertragung von Ergebnissen von einem Bereich auf den anderen zu beachten.

Üblicherweise ist der Einsatz eines Sichtsystems mit einer Absicht verknüpft, die man damit verfolgen will. Daher lautet der Titel der vorliegenden Arbeit „*Modellbasierte Lokalisation und Verfolgung für sichtsystemgestützte Regelungen*“. Meist ist mit der Bildauswertung – wie bei der sichtsystemgestützten Demontage – auch die Regelung von Aktuatoren verbunden. Weniger offensichtlich ist dies bei der Personenverfolgung. Selbst wenn in Zukunft lediglich ein Text generiert werden sollte, welcher die Personenbewegungen beschreibt, so ergibt sich eine Regelung der bildgebenden Kamera aus folgendem Grund. Da sich die Personen bewegen, sollten sie währenddessen stets von der Kamera erfaßt werden können. Installiert man die Kamera hinreichend weit ent-

fernt von der zu erfassenden Person, ergibt sich in der Regel ein kleines Personenabbild im auszuwertenden Bildmaterial. Dann ist allerdings zu erwarten, daß die Bewegungen nicht zuverlässig und genau genug erfaßt werden können. Es ist daher sinnvoll, die verwendete Kamera in einem geringeren Abstand zu installieren und der verfolgten Person programm geregelt nachzuführen. Das dokumentierte Verfahren liefert alle programmtechnischen Voraussetzungen, um eine solche Regelung einfach umsetzen zu können.

1.1 Personenverfolgung

Das in der vorliegenden Arbeit bei der Personenverfolgung eingesetzte System geht zurück auf [Wachter 97], siehe auch [Wachter & Nagel 99]. Das realisierte Programmsystem ermöglicht die Verfolgung, d.h. die Erfassung der Bewegungen von Personen in Bildfolgen realer Szenen über die Zeit hinweg. Dazu werden weder besondere Bewegungsmuster der Personen vorausgesetzt (insbesondere wird absichtlich kein Modell über den Gehvorgang eines Menschen eingesetzt) noch zu stark vereinfachende Bedingungen an die Kleidung oder den Hintergrund gestellt. Angestrebt wird somit die Entwicklung eines sehr flexiblen Verfahrens zur Operation in Realwelt-Szenen. Dies stellt allerdings sehr hohe Anforderungen an die eingesetzten Algorithmen. [Wachter 97], [Wachter & Nagel 99] gelingt es, diesen Anforderungen soweit gerecht zu werden, daß mehrere Personen in realen Szenen über mehrere hundert Bilder hinweg bis auf einige Robustheitsschwächen verfolgt werden können.

Das System operiert im 3D-Szenenbereich. Zur Modellierung der Körperteile werden elliptische Kegelstümpfe eingesetzt und über Koordinatensystemtransformationen bezüglich der zugehörigen lokalen Koordinatensysteme zu einem Personenmodell entsprechender Topologie verknüpft. Die Freiheitsgrade bei den Transformationen bezeichnen entweder personenspezifische Größen (z.B. Arm- oder Beinlängen) oder aus dem Bildmaterial zu schätzende Freiheitsgrade (wie beispielsweise die Beugewinkel der Knie oder der Ellenbogengelenke). Erstere können anhand der Person (bisher) manuell bestimmt werden. Letztere müssen für jedes ausgewertete Bild samt der zeitlichen Ableitungen (Geschwindigkeiten) aus dem Bildmaterial geschätzt werden. Zur Schätzung dieser insgesamt (je nach Anwendungsfall) bis zu 26 Freiheitsgrade wird ein geeigneter erweiterter Kalman-Filter eingesetzt zur Fortschreibung der Werte über die Bildfolge hinweg. Die Bewertung einer Schätzung trifft ein dabei zu minimierendes Gütemaß, die Minimierung erfolgt mittels des Gauß-Newton-Algorithmus. Dieses Gütemaß mißt in geeigneter Form die Abweichung zwischen dem ins Kamerabild projizierten Umriß des Personenmodells und den zugehörigen Bilddaten, wobei sowohl Kanteninformation als auch Flächeninformation des Bildes verwendet wird. Die Projektion des Personenmodellumrisses ins auszuwertende Kamerabild kann dabei aufgrund einer physikalisch-mathematischen Modellierung der Abbildungseigenschaften der verwendeten Kamera berechnet werden. Zur Visualisierung der Güte der Modellanpassung zu

einem Zeitpunkt der Verfolgung werden üblicherweise die zugrundeliegenden Bilddaten dargestellt und die zur Anpassung gehörige Projektion des Umrisses in Form weisser Linien überlagert.

Zur Stabilisierung der Verfolgung ist ein Mechanismus eingebaut, der es ermöglicht, die geschätzten Freiheitsgradwerte während der Modellanpassung zu beschränken. Dazu können Extremwerte für die zu schätzenden Freiheitsgrade festgelegt werden. Die Extremwerte können die anatomischen Beschränkungen der Gelenke des zu verfolgenden Objektes repräsentieren. Sie können aber auch allgemeiner so gewählt werden, daß sich möglichst gute Modellanpassungen ergeben. Falls das Programm einen bezüglich dieser Werte zu großen oder zu kleinen Gelenkwinkel berechnet, wird dieser auf den entsprechenden Extremwert und die zugehörige Änderungsgeschwindigkeit auf Null gesetzt. Dieser Mechanismus wird in der vorliegenden Arbeit mit *Freiheitsgradwertebeschränkung* bezeichnet.

In [Wachter 97] werden ausführlich die erhaltenen Verfolgungsergebnisse anhand von sechs Personenbildfolgen dokumentiert. Zu jeder der Bildfolgen wird bisher eine eigene Parametrisierung der Verfolgung verwendet. Abbildung 1.1 verschafft einen Überblick über die Verfolgungen. Die im oberen Teil der Abbildung dargestellten Bildfolgen zeigen gehende Personen vor Realwelthintergründen, bei der unten links dargestellten Teilabbildung vollführt die Person gymnastische Bewegungen und bei der unten rechts gezeigten Bildfolge wird eine Signalisierungsszene mit den Armen simuliert. Trotz der erzielten Resultate ergeben sich bei den Verfolgungen an einigen Stellen Schwachpunkte, auf die in [Wachter 97] nicht näher eingegangen werden konnte¹ und die es in der vorliegenden Arbeit zu verbessern gilt. Um auf die zum gegenwärtigen Stand erzielten Ergebnisse Bezug nehmen zu können, werden die Verfolgungsergebnisse nach [Wachter 97] in der vorliegenden Arbeit Referenzversuche genannt. Die Bezeichnungen der Bildfolgen sind der Bildunterschrift von Abbildung 1.1 zu entnehmen.

1.2 Sichtsystemgestützte Demontage

Bei der sichtsystemgestützten automatischen Demontage des Motorraums von Altfahrzeugen kann auf dem in [Tonko 97] entwickelten Verfahren aufgebaut werden. Als Versuchsumgebung steht eine Roboterzelle zur Verfügung, die mit zwei Robotern ausgestattet ist. Zwischen den Robotern befindet sich als Versuchsobjekt ein Altfahrzeug mit abgenommener Motorhaube, siehe Abbildung 1.2. Der eine Roboter trägt eine Stereokameraanordnung zur Aufnahme des benötigten Bildmaterials. Mit ihm können geeignete Beobachtungspositionen angefahren oder bestimmte Relativbewegungen zwischen Objekt und Kamera realisiert werden. Der werkzeugtragende Roboter (kurz Werkzeugroboter) kann – wie in der Abbildung dargestellt – einen Greifer tragen oder auch

¹In [Wachter 97] stand die Entwicklung eines funktionierenden Verfahrens im Vordergrund. An eine zeitaufwendige Analyse und Verbesserung der einzelnen Schwachstellen war zum damaligen Zeitpunkt (noch) nicht zu denken.



Abbildung 1.1: Referenzversuche: Verfolgung bei den sechs Bildfolgen nach [Wachter 97]. Bildfolgennamen und Halbbildzeitpunkte zur Bilderreihe oben, v.l.n.r.: `l_rot` (Halbbild 91), `alinde` (Halbbild 67), `frankm` (Halbb. 148) und `hof_191096` (Halbb. 129). Unten: `s1_191096` (Halbbild 47) (links) sowie `g_braun` (Halbbild 100) (rechts).



Abbildung 1.2: Versuchsumgebung Roboterzelle. Man kann den Beobachtungsroboter auf der linken Seite erkennen und den werkzeugtragenden Roboter im hinteren rechten Teil des Bildes.

mit anderen Werkzeugen ausgestattet werden, wie z.B. einem elektrischen Schrauber zum Lösen von (nicht zu festen) Schraubverbindungen. Der Blick in den Motorraum läßt die Herausforderungen an das zu entwickelnde Regelungssystem erahnen.

Als Steuerrechner diente seit [Tonko 97] die sogenannte GIGAmachine, ein Multiprozessorsystem aus mehreren Sparc 10 Prozessoren zu maximal 75 MHz. Zur genaueren Erläuterung siehe [Tonko 97].

In [Gengenbach 94] speziell entwickelte Geräte, genannt MiniVISTAs, übernehmen den Bildeinzug von den Kameras und die Übertragung an den Steuerrechner sowie die Echtzeit-Extraktion der bei der Verfolgung benötigten Kantenelemente. Diese Anordnung liefert mit einem Takt von unter 40 ms Halbbilder inklusive der zugehörigen Kantenelementbilder und ermöglicht so überhaupt erst, in den Bereich einer Echtzeitregelung kommen zu können, da die Berechnung der Kantenelementbilder per Programm sehr rechenaufwendig ist. Das Kantenelementbild beinhaltet diejenigen Pixel im Bild, die als Bildkantenelemente identifiziert worden sind (d.h. deren Gradientenbetrag über einer fest vorgegebenen Schwelle liegt *und* deren Gradientenbetrag in Gradientenrichtung ein Maximum annimmt) samt der dort jeweils bestehenden Gradientenrichtung. Abbildung 1.3 zeigt zur Illustration ein Beispiel.

Zur Verfolgung der Objekte werden diese mittels polyedrischer CAD-Modelle repräsentiert und Rundungen polyedrisch approximiert. Die Objekte können monokular oder binokular verfolgt werden. Ähnlich wie bei der Personenverfolgung dient auch hier die Projektion des Modells in die Bildebene der Berechnung einer zu minimierenden Anpassungsgüte sowie der Beurteilung der berechneten Modellanpassung. Anders als bei der



Abbildung 1.3: Kantenelementbild (unten) zu einer Aufnahme der Autobatterie (oben). Die Gradientenrichtungen der Kantenelemente sind in den jeweils gesetzten Pixeln kodiert. Aus Zeitgründen werden bei der Verfolgung stets Halbbilder, keine Vollbilder verwendet.

Personenverfolgung werden aus Zeitgründen ausschließlich Kanteninformationen, keine Flächeninformationen zur Anpassung genutzt. Der Lagebestimmung von Objekten dient ein erweiterter Kalman-Filter mit Gauß-Newton-Minimierung der Gütefunktion. Da die Lage eines Objektes im Raum durch 6 Parameter definiert ist (drei Translationen und drei Rotationen) müssen je Objekt 6 Freiheitsgradwerte über die Zeit hinweg fortgeschrieben werden. Die Initialisierung der Verfolgung geschieht in [Tonko 97] manuell.

Das in [Tonko 97] entwickelte System ist in der Lage, die Batterie oder den Luftfilter bei sich bewegenden Kameras mono- oder binokular zu verfolgen. Zur Illustration sind in Abbildung 1.4 die monokulare Verfolgung der Autobatterie samt der im Anpassungsprozeß verwendeten Kantenelemente gezeigt und in Abbildung 1.5 die Stereoverfolgung des Luftfilters. Als Demontagebeispiel dient die sichtsystemgestützte Entnahme der (von den Halterungen bereits befreiten) Autobatterie aus dem Motorraum, vgl. Abb. 1.6. Zur Regelung des Greifers an die korrekte Greifposition wird simultan sowohl die Batterie als auch der Greifer binokular verfolgt. Während des gesamten Prozesses werden aus der Relativlage der beiden Objekte zueinander entsprechende Stellbefehle für den Werkzeugroboter generiert. Auf diese Weise wird weder eine externe Kamerakalibrierung noch eine Roboterkalibrierung benötigt (welche oft mit nicht vernachlässigbaren Fehlern behaftet sind), und eventuelle Ungenauigkeiten werden während der Regelung automatisch ausgeglichen.

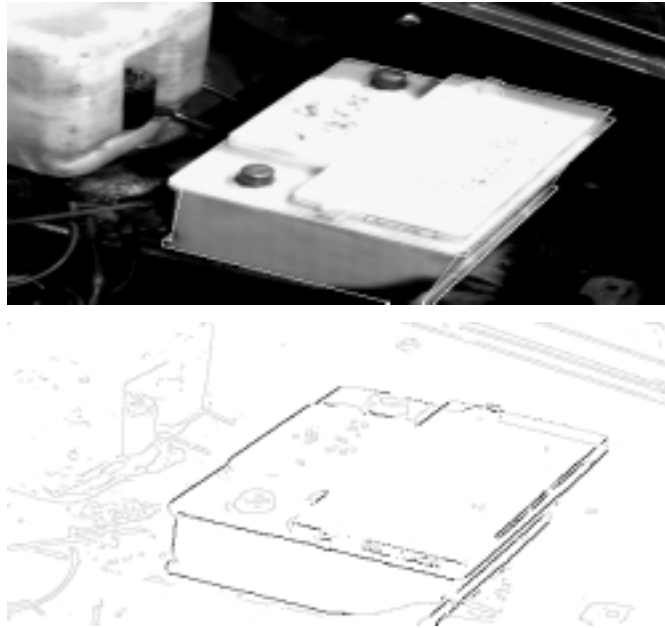


Abbildung 1.4: Monokulare Verfolgung der Autobatterie zum Halbbildzeitpunkt 3 (oben). Der untere Teil zeigt das zugehörige Kantenmodell. Die im Anpassungsprozeß verwendeten Kantenelemente wurden gegenüber den nicht verwendeten schwarz hervorgehoben.

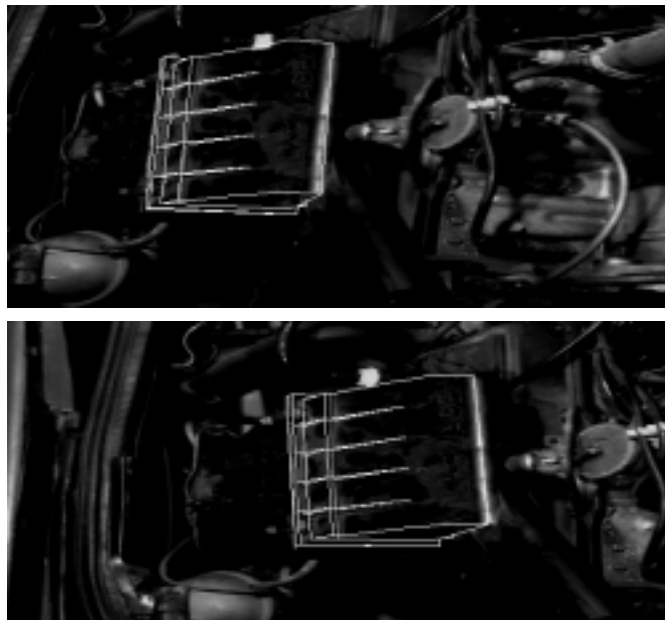


Abbildung 1.5: Stereoverfolgung des Luftfilters: Projektion des Modells ins linke Kamerabild (oben) bzw. ins rechte (unten).

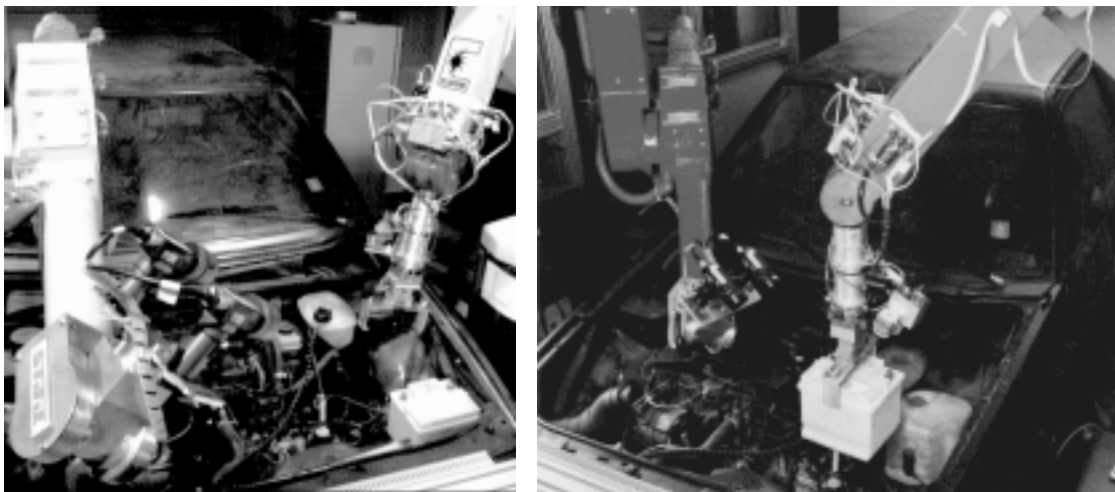


Abbildung 1.6: Regelung des Greifers an die korrekte Greifposition (links) und Entnahme der Batterie aus dem Motorraum (rechts).

Mit dem eben umrissenen System konnte gezeigt werden, daß es prinzipiell möglich ist, geeignete Regelungen für den Motorraum zu realisieren. Damit liegt eine Basis vor für die hier folgenden Betrachtungen.

1.3 Ziel der vorliegenden Arbeit

Ziel der vorliegenden Arbeit ist eine signifikante Verbesserung hinsichtlich der in Theorie und Praxis relevanten Aspekte, wie z.B. Automatisierbarkeit, Robustheit, Genauigkeit, Flexibilität, Ausführungsgeschwindigkeit, Zuverlässigkeit und Analysierbarkeit. Diese Aspekte treten in natürlicher Weise bei beiden Teilbereichen während der Weiterentwicklung mit unterschiedlicher Gewichtung unter unterschiedlichen Rahmenbedingungen auf. Dieser Sachverhalt ermöglicht besonders gut die Darstellung der bei modellbasierten Sichtsystemen relevanten Aspekte und die zu deren Verbesserung notwendigen Maßnahmen. Gleichzeitig wird aufgrund der Verschiedenartigkeit der beiden Teilbereiche eine breite Abdeckung des Feldes modellbasierter Sichtsysteme erzielt.

Die Verbesserung des Systems hinsichtlich oben genannter Bewertungsaspekte erfordert auf der einen Seite die Entwicklung neuer Verfahren. Auf der anderen Seite ist aber auch eine Analyse und Verbesserung zahlreicher Einflußfaktoren unumgänglich, wie sich zum Beispiel insbesondere bei der Steigerung der Robustheit zeigt. Es ergeben sich somit zwei Möglichkeiten des Fortschritts. Beide werden in der vorliegenden Arbeit behandelt und jeweils signifikante Fortschritte im Hinblick auf die genannten Bewertungsaspekte dokumentiert. Obgleich die Verbesserung zahlreicher Einflußfaktoren in einigen Fällen mit zum Teil sehr einfachen Maßnahmen bewerkstelligt werden kann, ist deren Gesamtwirkung auf das System kaum zu unterschätzen und daher von großer

Bedeutung für die Verbesserung des Systems.

Die Kapitel 3 und 4 dokumentieren die konkret auf den beiden Teilgebieten erzielten Ergebnisse und wie die zugrundeliegenden Bewertungsaspekte im Rahmen der Weiterentwicklung des Systems in natürlicher Art und Weise entstehen. Dabei wird verdeutlicht, welche Ebenen der Entwicklung zum Tragen kommen können. Zur Vertiefung der bei komplexen Systemen während der Entwicklung auftretenden Faktoren bzw. Möglichkeiten sei auf den sehr interessanten Artikel [Rechtin 97] hingewiesen.

Zur Vervollständigung der Untersuchungen werden neben den erzielten Fortschritten auch die noch bestehenden Restschwierigkeiten aufgezeigt, für die noch keine Lösung gefunden werden konnte. Diesbezüglich bereits gewonnene Erkenntnisse mögen als Grundlage für zukünftige Verbesserungen dienen.

Kapitel 2

Literaturüberblick

2.1 Erkennung und Lokalisation

Die anschließende Diskussion der Literatur zur Erkennung und Lokalisation von Objekten aus 2D-Einzelaufnahmen soll dem Leser ermöglichen, einen Überblick über die auf diesem Sektor entwickelten Verfahren und deren prinzipielle Unterschiede zu erlangen. Dazu ist als erstes die in der Literatur und auch im Folgenden verwendete Terminologie zu klären. Konkret sind folgende drei Begriffe gegeneinander abzugrenzen:

- Mit dem Begriff der *Detektion* bezeichnet man die algorithmische Feststellung, daß (irgend) ein Objekt in der vorhandenen Aufnahme sichtbar (geworden) ist. Ein dazu herangezogenes Verfahren erkennt nicht notwendigerweise, um welches Objekt oder um welchen Objekttypen es sich tatsächlich handelt. Um dies zu veranschaulichen, sei an einen Algorithmus gedacht, der in einer Bildfolge den Ausschnitt eines Bildes und den gleichen Ausschnitt einer Referenzaufnahme voneinander subtrahiert und durch Vergleich des Differenzbetrages mit einer geeigneten Schwelle feststellen kann, ob an dieser Stelle etwas im Bild aufgetaucht ist oder nicht.
- *Erkennung* bezeichnet die Identifikation einer Auswahl von Bildstrukturen als Abbild eines bestimmten Objektes einer Datenbasis. Es wird also festgestellt, *welches* Objekt im Bild sichtbar ist.
- Der Begriff der *Lokalisation* bezeichnet schließlich die Berechnung der Lage und Orientierung eines Objektes in Bild bzw. Szene aus den vorhandenen Bildstrukturen.

Gemäß der Begriffsbildung sind die genannten Aufgaben zur Lokalisation von Objekten sequentiell in der dargestellten Reihenfolge durchzuführen. Tatsächlich hängen diese drei Begriffe jedoch je nach eingesetztem Verfahren unterschiedlich stark zusammen. Viele Verfahren lösen beispielsweise die Aufgabe der Erkennung und Lokalisation

untrennbar voneinander in einem Algorithmus. Es gibt aber auch Verfahren, welche beispielsweise eine Erkennung realisieren und erst nach Durchführung einer nachgeschalteten Verarbeitung der dabei anfallenden Daten eine Lokalisation bewerkstelligen.

Wie die Autoren in [Bebis et al. 98] feststellen, ist die Objekterkennung eine schwierige Angelegenheit: 'Recognizing 3D objects from 2D images is one of the most challenging problems in computer vision'. Dies spiegelt sich darin wider, daß mittlerweile viele Ansätze von zum Teil sehr unterschiedlicher Natur entwickelt und weiterentwickelt wurden. Und immer noch scheint kein Ende dieser Entwicklung in Sicht zu sein. Die bis 1989 veröffentlichten Verfahren werden abgedeckt in dem empfehlenswerten Buch [Grimson 90]. Darin findet sich neben der Definition der sich stellenden Aufgabe und zahlreichen Illustrationen anhand anschaulicher Beispiele eine systematische Zusammenstellung der bis 1989 bekannten Ansätze und der sich stellenden Teilprobleme. Es werden unter anderem behandelt: modellbasierte Ansätze, Zuordnung zwischen Bild- und Modellmerkmalen, Suchmethoden im Raum möglicher Objektlagen sowie Suchmethoden im Raum möglicher Zuordnungen, Verfahren zur Reduzierung der Komplexität oder der Dimensionalität des Suchraumes, Abtastmethoden, Verfahren mittels (generalisierter) Hough-Transformation und die Hypothesenverifikation. [Wheeler & Ikeuchi 95] geben eine ausführliche Übersicht über die bis einschließlich 1994 veröffentlichten Verfahren. Für Arbeiten vor 1995 sei daher auf diese Übersicht verwiesen. Als Ergänzung ist die Literaturübersicht in [Olson 99] zu erwähnen, in der einige wichtige Verfahren neueren wie auch älteren Datums (wie z.B. die in der Fachwelt oft zitierte Methode 'RANSAC') angesprochen werden. Aufgrund der durch die genannten Zusammenfassungen erreichte Abdeckung bis (mindestens) 1994 konzentriert sich die folgende Literaturrecherche auf Artikel seit 1995, wobei durch neuere Artikel überholte Arbeiten nicht mit aufgenommen sind.

Da die meisten wichtigen Arbeiten auf dem Gebiet der Bildauswertung in der (meist) jährlich erscheinenden Übersicht von A. Rosenfeld in der wissenschaftlichen Zeitschrift 'Computer Vision and Image Understanding' zitiert werden, bilden die dort berücksichtigten Zeitschriften und Konferenzbände die Hauptgrundlage des hier durchsuchten Literaturbestandes. Berücksichtigt wurde ferner der Bestand in den Bibliotheken der Universität Karlsruhe sowie die im Internet zugänglichen Arbeiten.

Im Folgenden wurde ein Versuch unternommen, die Fülle an unterschiedlichen Artikeln zur Objekterkennung bzw. -lokalisierung zu klassifizieren und die einzelnen Ansätze zu umreißen. Bei der Klassifikation wurde dabei keine disjunkte Einteilung angestrebt, sondern eine zur übersichtlichen Darstellung möglichst geeignete. Mischformen zeigen, daß die Grenzen zwischen den gewählten Klassen fließend sein können.

Da naturgemäß verschiedene Verfahren i.A. unterschiedliche Stärken und Schwächen aufweisen, wurden zu den einzelnen Artikeln die aufgeführten Testbeispiele genannt. Solche Beispiele vermitteln meist einen guten Eindruck von den Stärken und Schwerpunkten, dem Einsatzgebiet bzw. der Realwelttauglichkeit der Verfahren.

2.1.1 Merkmalsbasierte Verfahren

Merkmalsbasierte Verfahren beruhen auf der *Extraktion von markanten Merkmalen in Modell und Bild* sowie einer geeigneten *Zuordnung* zwischen den Modell- und Bildmerkmalen. Mit Hilfe der Bild-Modell-Zuordnung, die im Bildbereich oder einem geeigneten Transformationsraum durchgeführt wird, kann auf das Vorhandensein eines bestimmten Objekts im Bild oder auf dessen Lage geschlossen werden. Die verwendeten Merkmale können z.B. Punkte, Linien, Eckstrukturen, Ellipsen oder von allgemeinerer, evtl. geeignet transformierter Art sein. Typischerweise werden die berechneten Lagehypothesen anschließend einem Verifikationsschritt unterzogen. Aufgrund der Funktionsweise solcher Verfahren ergeben sich direkt die zu lösenden zentralen Probleme der merkmalsbasierten Objekterkennung bzw. -lokalisierung:

- Auswahl von Merkmalen sowie
- deren zuverlässige Extraktion aus Bildern,
- Problem der Merkmalszuordnungen und die damit verbundene kombinatorische Vielfalt (Problem der Komplexität bei der Zuordnung),
- Generierung zuverlässiger, möglichst genauer Lagehypothesen; insbesondere die Behandlung von Ausreißern (d.h. Fehlzusordnungen) spielt hierbei eine wichtige Rolle,
- Realisierung der Verifikation.

Diese Punkte spiegeln sich direkt in den Stoßrichtungen der meisten Verfahren wider. So beschäftigen sich die meisten der im folgenden zitierten Artikel mit der Lösung dieser Teilprobleme und der diesbezüglichen Verbesserung oder Optimierung bestehender Verfahren. [Yang & Cohen 99] faßt einige wichtige Stoßrichtungen merkmalsbasierter Verfahren mittels entsprechender Literaturzitate zusammen.

[Gengenbach et al. 96] verwenden für die Lokalisation polygonaler Körper einen klassischen merkmalsbasierten Ansatz [Porill et al. 88] zur Generierung und Verifikation hypothetischer Objektlagen. Als Merkmale werden sowohl Modellecken als auch Modellkreise bzw. Bildecken und Bildellipsen sowie bei der Verifikation auch geradlinige Kanten verwendet. Die Autoren nutzen eine trinokulare Kameraanordnung, um aus den 2D-Bildinformationen möglichst zuverlässige und eindeutige 3D-Rekonstruktionen zu erhalten und so die Komplexität des Zuordnungsproblems eingrenzen sowie Fehlzusordnungen möglichst ausschließen zu können. Als Testbeispiel dient ein aus zwei Quadern unterschiedlicher Größe zusammengesetzter Körper vor homogenem Hintergrund.

Der in [Cass 98] vorgestellte Ansatz basiert auf einer Projektion geometrischer Zwangsbedingungen, die bei der Zuordnung geometrischer Merkmale in Bild und Modell bestehen, in einen geeigneten Transformationsraum. Die Analyse der Zwangsbedingungen in

diesem Transformationsraum erlaubt es, ein Zuordnungsverfahren anzugeben, welches alle möglichen Zuordnungen zwischen Modell- und Bildmerkmalen mit polynomialem Zeitaufwand findet. Getestet wird das Verfahren anhand eines aus zwei Buchstützen zusammengeklebten Objektes in freier Lage vor homogenem Hintergrund.

[Chang & Tsai 99] stellen einen grundlegenden, auf Bildkantenmerkmalen basierenden Ansatz vor, der mit statistischen Tests der gewonnenen Lagehypothesen arbeitet. Dazu werden zwei Testfunktionen definiert. Mit der ersten kann ungünstiges Bildmaterial noch vor der eigentlichen Lageschätzung erkannt werden. Mit der zweiten wird entschieden, ob die geschätzte Lage genau genug berechnet werden konnte. Als Testbeispiele dienen der Zauberwürfel (auch bekannt als ‘Rubik’s Cube’) und ein Drucker vor homogenem Hintergrund.

[Yang & Cohen 99] wenden sich nicht der Lagebestimmung dreidimensionaler Objekte im Raum zu, sondern lokalisieren eine zweidimensionale Bildschablone, in der schwache perspektivische Transformationen der Schablone auftreten dürfen, im Abbild einer Szene. Solche perspektivische Transformationen werden mittels affiner Transformationen approximiert. Die Bildschablonen können beispielsweise Aufkleber auf 3D-Objekten oder Oberflächenteile repräsentieren und so indirekt eine 3D-Lokalisation ermöglichen. Den Kern des vorgestellten Verfahrens bildet die Berechnung und Nutzung lokaler affiner Invarianten. Deren Berechnung baut auf der Extraktion von Punktmerkmalen im Bildbereich und anschließender Bestimmung der konvexen Hülle auf. Durch Betrachtung konsekutiver Eckpunkte der konvexen Hülle und der daraus konstruierten Teildreiecke werden die affinen Invarianten bestimmt. Da diese affinen Invarianten lokal konstruiert werden, können partielle Verdeckungen der Objekte direkt berücksichtigt werden. Die affinen Invarianten werden zur Berechnung von Zuordnungen zwischen den Knoten der konvexen Hüllen von Testbild und Schablone genutzt, um die affine Transformation zwischen Testbild und Schablone rückgängig machen zu können. Das eigentliche Erkennungs- bzw. Lokalisationsproblem löst ein polynomialer Algorithmus zur Zuordnung der Punktmerkmale in Testbild und Schablone. Als Testbeispiel dient zum einen der Schriftzug auf einer Coladose (‘Gotta have it. Pepsi’), der in mehreren Realweltszenen lokalisiert wird. Zum anderen wird aber auch ein Haus in mehreren Bildern lokalisiert, die verschiedene Ansichten des betrachteten Hauses zeigen.

Der in [Lowe 99] vorgestellte Ansatz basiert ebenfalls auf lokalen Invarianten. Die extrahierten speziellen Merkmale sind invariant gegenüber Skalierung, Translation sowie Rotation und teilweise invariant gegenüber Beleuchtungsänderungen sowie affinen bzw. 3D Projektionen. Die Merkmale weisen laut Autor ähnliche Eigenschaften auf wie die grundlegenden, zur Objekterkennung genutzten Neuronen der Hirnrinde bei Primaten. Durch die Verwendung von speziellen Bildschlüsseln können lokale geometrische Deformationen eingefangen werden, indem verwischte Gradientenbilder in mehreren Orientierungs- und Skalierungsstufen repräsentiert werden. Die Schlüssel dienen der Bestimmung von möglichen Kandidaten für eine Objektanpassung. Diese werden in einem abschließenden Verifikationsschritt mittels einer geeigneten Bewertungsfunktion geprüft. Die im Artikel angegebenen Testbeispiele überzeugen durch ihre hohe

Komplexität: unter zum Teil gravierenden partiellen Verdeckungen werden beschriftete Verpackungsschachteln, ein Sportschuh, ein Telefon sowie ein Teddybär vor hochkomplexem Realwelthintergrund lokalisiert. Abschließend demonstriert der Autor die Stabilität der Bildschlüssel bei Beleuchtungsänderung. Es sollte erwähnt werden, daß das Verfahren eine Parallele zu den betrachtungswinkelbasierten Verfahren aufweist. Die Abhängigkeit vom Betrachtungswinkel ist allerdings wesentlich geringer als bei den in Abschnitt 2.1.8 behandelten Verfahren.

Auch [Weiss & Ray 98] arbeiten mit Invarianten. Unter einigen Annahmen zur Modellierung der Objekte finden die Autoren mehrere Invarianten bezüglich der Projektion von 3D nach 2D, welche die Grundlage des vorgestellten Ansatzes bilden. Die Annahmen zur Modellierung können dabei ein spezielles Modell oder eine Klasse von Modellen betreffen. Beide Fälle werden behandelt. Durch eine Behandlung des Zuordnungsproblems zwischen Modell- und Bildmerkmalen gelingt es den Autoren, Gelände-PKW trotz schwachem Kontrast und z.T. starken Schattenbildungen auf einer Wiese mit einfacher Textur zu lokalisieren.

[Das & Bhanu 98] stellen sich der Aufgabe, Flugzeuge unter Realweltbedingungen zu detektieren, zu erkennen und zu lokalisieren. Wie die Testbeispiele zeigen, ist die gestellte Aufgabe aufgrund der Form mancher Flugzeuge, wegen hochkomplexem Hintergrund, Schattenwurf sowie z.T. schwachem Kontrast sehr komplex. Um die Aufgabe zu lösen, verwenden die Autoren eine hierarchisch organisierte Datenbasis aus Objektmodellen und stellen einen Algorithmus vor, den sie in drei Teile untergliedern. Den ersten Teil bildet ein Verfahren zur qualitativen Objekterkennung unter Verwendung einer modellbasierten, symbolischen Merkmalsextraktion. Der zweite Teil führt eine verfeinerte Anpassung und eine spezifischere Klassifikation durch, indem verfeinerte Merkmale verwendet werden. Dabei werden Eingabedaten aus dem ersten und dritten Teil verwendet. Der dritte Teil extrahiert Basismerkmale, gewichtet diese geeignet anhand ihrer Wichtigkeit und beeinflußt damit die beiden anderen Teile. Die Funktionstüchtigkeit des beschriebenen Algorithmus wird anhand vieler Flugzeugtypen unter z.T. schwierigsten Bedingungen belegt. Die Realwelttauglichkeit des Verfahrens wird damit eindrucksvoll unter Beweis gestellt.

[Rosin 99] widmet sich der Tatsache, daß die auf der Lösung der *kleinsten Quadrate* basierenden Standardmethoden zur Berechnung des Zuordnungsproblems sehr empfindlich auf Ausreißer in Form von Fehlzuordnungen reagieren. Der Autor setzt zur Robustheitssteigerung an diesem Punkt an und stellt einen sich auf Punktzuordnungen stützenden Ansatz vor, welcher mit dem *kleinsten Median der Quadrate* arbeitet. Der vorgestellte Algorithmus läuft mit linearem Aufwand in der Anzahl der Punktmerkmale und liefert bis zu einem Anteil von 50% Ausreißern robuste Lageschätzungen. Die Robustheit des Verfahrens wird anhand zweier Arten von Störungen in den Eingabedaten getestet: zum einen unter geringer Gaußscher Störung aller Daten und zum zweiten unter massiver Störung weniger Eingabedaten. Als Testobjekt bei der Lagebestimmung dient ein einfaches polyedrisches Werkstück vor einfach strukturiertem Hintergrund.

[Han et al. 99] wenden sich dem zentralen Zuordnungsproblem bei der merkmalsbasierten Objekterkennung zu. Die Autoren präsentieren einen Algorithmus für die Zuordnung zwischen den 3D Modell- und den 2D Bildmerkmalen, der im Gegensatz zu den herkömmlichen Ansätzen auf einer Merkmalsmenge *variabler* Kardinalität operiert und nicht ein einzelnes lokales Merkmal oder eine fixe Merkmalskardinalität voraussetzt. Als Maß für die Ähnlichkeit zwischen Merkmalsmengen dient die Hausdorff-Distanz¹ (vgl. [Huttenlocher et al. 93]) mit expliziten Zuordnungen, die von den Autoren geeignet erweitert wird, um die Probleme im Zusammenhang mit im Bild nicht vorhandenen oder stark von der Erwartung abweichenden Merkmalen lösen zu können. Nach der Generierung der Lagehypothesen erfolgt - wie üblich - ein Verifikationsschritt. Als Testobjekte dienen den Autoren partiell verdeckte Polyeder in komplexer Umgebung.

[Häusler & Ritter 99] bestimmen die sechs Lage- und Orientierungsfreiheitsgrade eines zu lokalisierenden starren Objekts, indem zunächst aus dem Bildmaterial Punktmerkmale extrahiert und aggregiert werden. Anschließend wird jede dieser so gewonnenen Aggregationen mit einer internen Datenbasis aus entsprechenden Merkmalsrepräsentationen der 3D Objekte verglichen. Jedes Bildmerkmal kann auf diese Weise auf einen Teilraum möglicher Lösungen für die Lageparameter abgebildet werden. Mittels einer Hough-Tabelle können die so gewonnenen Lagehypothesen akkumuliert und durch eine Suche nach dem Maximum in der Tabelle die Objektlage berechnet werden. Durch die Verwendung der Hough-Tabelle werden automatisch partielle Verdeckungen der Objekte toleriert. Mit dem vorgestellten Verfahren werden aus Quadern und Kegeln unterschiedlicher Größe zusammengesetzte mechanische Bauteile sowie eine Beethoven-Büste, d.h. eine Freiformskulptur, vor homogenem Hintergrund lokalisiert.

In ähnlicher Weise wie der vorangehend behandelte Artikel bildet [Jurie 98b] mögliche Anpassungen zwischen Bild- und Modellmerkmalen auf Teilräume eines geeigneten Oberraumes ab. Der Autor verwendet dafür bestimmte Volumina in einem Transformationsraum, von denen anschließend zur Lösung der Lokalisationsaufgabe dasjenige bestimmt wird, welchem die größte Anzahl an Korrespondenzen zwischen Bild- und Modellmerkmalen zugeordnet worden sind. Mit dem beschriebenen Verfahren wird eine Computermaus vor komplexem Hintergrund lokalisiert und unter z.T. signifikanten partiellen Verdeckungen (bei alleiniger Verwendung der Lokalisation) erfolgreich über die Zeit hinweg verfolgt.

[Mirmehdi et al. 99] stellen Strategien vor zur effizienten Kombination von Hypothesengenerierung und Verifikation, indem sie beide Teile in eine Kontrollschleife einbinden. Die Autoren verbessern die klassischen Bildauswertungstechniken, indem sie Kontrollmechanismen auf drei unterschiedlichen Ebenen der Analyse angeben. In der Schleife werden bestimmte Mengen an Merkmalen aus dem Bild extrahiert, um die Anzahl

¹Die (gerichtete) *Hausdorff-Distanz* zwischen zwei Mengen A und B ist definiert als

$$h(A, B) := \max_{a \in A} \min_{b \in B} \|a - b\| ,$$

wobei $\|\cdot\|$ eine beliebige Norm wie z.B. die euklidische Norm bezeichnet.

an zu verifizierenden Lagehypothesen zu reduzieren. Die durch die Schleife realisierte Rückkopplung erlaubt dabei eine sukzessive Verbesserung anhand der Verifikation der Merkmalsmengen, sodaß die Objekte sogar in sehr verrauschtem Datenmaterial lokalisiert werden können. Wenn zu einem lokalisierten Objekt zusätzliche Informationen aus dem Bild extrahiert werden sollen, erlaubt der vorgestellte Mechanismus eine Konzentration der Suche auf die fehlenden Informationen. Der konkret von den Autoren umgesetzte Algorithmus lokalisiert auch partiell verdeckte Objekte in einer rotations- und skalierungsunabhängigen Weise. Als Testbeispiele werden einfache Quader in einer Klötzchenwelt, eine partiell verdeckte, beschriftete Schachtel sowie verschiedene, von einem Flugzeug aus aufgenommene Straßenbrücken in Realweltszenen lokalisiert.

In [Jurie 98a] und [Jurie 99] wendet sich der Autor dem neben dem Zuordnungsproblem zweiten zentralen Problem der merkmalsbasierten Erkennung zu: der Verifikation der Lagehypothesen. Er stellt fest, daß dieser Schritt besonders bei komplexen Szenen bzw. verrauschten Eingabedaten ein schwieriges Problem darstellt, da die generierten Hypothesen oft zu ungenauen Anpassungen zwischen Modell(projektion) und Objektabbild führen. Unter Zugrundelegung eines Gaußschen Modells für den Fehler in den Eingabedaten stellt der Autor einen robusten Verifikationsalgorithmus vor, der solche ungenauen Anpassungen berücksichtigt bzw. korrigieren kann, und der bei 2D und 3D Erkennungsaufgaben eingesetzt werden kann. Die Korrektur der ungenauen Anpassungen geschieht dabei durch sukzessive Unterteilung des Raumes möglicher Objektlagen und Bewertung der Teilräume, bis der verbleibende Teilraum hinreichend geschrumpft ist. Der Algorithmus wird u.a. zur 2D-Lokalisation eines Spielzeugpferdes, eines Spielzeugautos, einer Diskette und einer Computermaus sowie zur 3D-Lokalisation der Computermaus vor komplexem Hintergrund demonstriert, wobei zum Teil partielle Verdeckungen auftreten. Die hierbei eingesetzte 3D-Lokalisation funktioniert betrachtungswinkelbasiert (vgl. Abschnitt 2.1.8).

[Lu et al. 00] weisen auf eine prinzipielle Schwäche bei den nach Ansicht der Autoren besten merkmalsbasierten Verfahren hin. Zur Lageschätzung aus Punktmerkmalen des Bildbereichs verwenden diese Verfahren iterative Optimierungsmethoden, deren Konvergenz im Allgemeinen nicht bewiesen werden kann oder welche die Orthonormalitätseigenschaft der zu bestimmenden Rotationsmatrizen nicht zufriedenstellend berücksichtigen. Um diese Schwäche zu beseitigen, stellen die Autoren einen iterativen Algorithmus zur Lageschätzung aus Punktmerkmalen vor, welcher auf der Minimierung eines speziellen Fehlermaßes basiert und direkt orthonormale Matrizen für die Rotation liefert. Die Autoren betonen, daß der Algorithmus global konvergiert, effizient arbeitet, mindestens so genaue Ergebnisse berechnet wie die betrachteten Referenzverfahren und diese hinsichtlich der Robustheit gegenüber Ausreißern sogar übertrifft.

[Joseph 99] greift die bei der merkmalsbasierten Objekterkennung oft eingesetzten erschöpfenden Algorithmen zur Bestimmung der Zuordnungen zwischen den Modell- und Bildmerkmalen auf. Da diese Algorithmen zuverlässig funktionieren, jedoch sehr rechenzeitintensiv sind, wird der Rechenaufwand genauer analysiert und kann in Spezialfällen deutlich reduziert werden.

Abschließend sei auf eine interessante, umfassende Abhandlung [Alter & Jacobs 98] über die Fehlerpropagierung bei der merkmalsbasierten Lokalisation hingewiesen. Es wird dargestellt, wie unvermeidliche Fehler in den Eingabedaten sich auf die gefundenen Objektlagen niederschlagen. Dies verdeutlicht abschließend nochmals den zentralen Problemkreis bei der merkmalsbasierten Lokalisation und läßt einzelne Stoßrichtungen obiger Artikel noch klarer werden (vgl. insbesondere [Jurie 98a] und [Jurie 99]).

2.1.2 Freiformkurvenbasierte Verfahren

Die in diesem Abschnitt vorgestellten Verfahren zur Objekterkennung und -lokalisierung basieren auf der Analyse des durch *Freiformkurven* bestimmten *Objektumrisses* der betrachteten Objekte im Bild.

[Paulik & Wang 98] lösen das Problem, verschiedene Flugzeuge am Boden und in der Luft anhand ihres Umrißes erkennen bzw. lokalisieren zu können, indem sie die im Bild extrahierten, krummlinigen Objektumrisse einer Vektor-Wavelet-Analyse unterziehen. Darauf aufbauend entwickeln die Autoren eine 3D-Lokalisierung, deren zugrundeliegender Modellierungsteil invariant ist gegenüber Skalierung, Verschiebung, Rotation und dem Startpunkt der Umrißlinie. Es versteht sich fast von selbst, daß der Ansatz die Berücksichtigung des Betrachtungswinkels erfordert (vgl. Abschnitt 2.1.8). Das Verfahren wird angewandt zur Lokalisation von acht verschiedenen Flugzeugen in verschiedenen Lagen am Boden und in der Luft. Die Lokalisation verläuft bei einer Testdatensmenge von über 700 Flugzeugbildern je nach Versuchsreihe in 91,5% bis 96,8% der Läufe erfolgreich.

Das in [Vijayakumar et al. 98] vorgestellte Verfahren zur Erkennung dreidimensionaler Objekte anhand ihrer glatten, gekrümmten Silhouette im Bild basiert auf einer betrachtungswinkelunabhängigen Beziehung zwischen der Objektgeometrie und bestimmten Bildmerkmalen unter schwacher perspektivischer Projektion. Die Bildmerkmale selbst sind betrachtungswinkelabhängig. Konkret bildet die Menge aller zweimalig die Objektsilhouette berührenden Tangenten zusammen mit allen Konturpunkten gleicher Tangentialrichtung die Projektion einer eindimensionalen Menge von Oberflächenpunkten. Dabei liegt jeder dieser Oberflächenpunkte auf der Objektkontur, welche letztendlich von einer durch fünf Parameter beschriebenen Familie von Betrachtungswinkeln abhängt. Die so gewonnenen Bildmerkmale bilden damit eine einparametrische Familie von Äquivalenzklassen. Es wird gezeigt, daß jede dieser Klassen durch eine Menge numerischer Attribute charakterisiert werden kann, die über der korrespondierenden 5-dimensionalen Menge von Betrachtungswinkeln konstant bleibt. Auf diese Weise können die zu erkennenden Objekte durch Umrißkurven beschrieben werden, die in hochdimensionalen Räumen eingebettet sind. Die Modellierung geschieht *ohne* Kenntnis der Transformation zwischen Objekt und Kamera durch Aufnahme des Objekts mittels der Kamera. Das bedeutet insbesondere, daß keine implizite oder explizite 3D Formenrekonstruktion zum Einsatz kommt. Zur Elimination potentieller falscher

Anpassungen werden quantitative und qualitative Verifikationsprozeduren eingesetzt. Unter Anwesenheit von z.T. mehreren Objekten im Bild werden eine Stockente unter verschiedenen Betrachtungswinkeln sowie eine Vase bzw. ein rotationssymmetrisches Teil einer Inneneinrichtung (vor homogenem Hintergrund) lokalisiert.

2.1.3 Freiformflächenbasierte Verfahren

Tatsächlich kann man auch Verfahren konstruieren, welche eine *Erkennung und Lokalisation von 3D-Freiformkörpern mittels* der Anpassung von *Freiformmodellen* ermöglichen, wie die Arbeit [Johnson & Hebert 99] zeigt. Zur Realisierung der Aufgabe führen die Autoren eine geeignete Repräsentation der Objektmodelle ein (engl.: ‘spin images’), welche aus polygonalen, Flächennormalen-behafteten Oberflächennetzen berechnet werden kann und an jedem Netzknoten die Dichte der Flächenpunkte abhängig von den Oberflächennormalen beschreibt. Auf diese Weise gelingt es den Autoren, die Anpassung von Freiformflächen durch die Anpassung von Punktmengen zu realisieren. Die Autoren lokalisieren mit dem vorgestellten Ansatz viele verschiedene, z.T. partiell verdeckte Freiformobjekte, wie z.B. eine Gummiente, andere Spielzeugtiere und -fahrzeuge oder diverse Rohrstücke in hochkomplexen Szenen.

Dieser Artikel bildet übrigens eine interessante Brücke zwischen den Disziplinen Bilderzeugung und Bildauswertung. Eine ausführliche Erörterung zu diesem Thema findet sich in [Lengyel 98].

2.1.4 Überführung von Punktmengen ineinander

Die in diesem Abschnitt beschriebenen Verfahren bilden eine spezielle Untermenge der merkmalsbasierten Verfahren. Sie realisieren eine *spezielle Überführung zweier Punktmengen aufeinander*. Dabei steht die eine Punktmenge z.B. für extrahierte Silhouettenpunkte oder geeignete Innenpunkte eines Objektes im Bild, die andere für ein Modell der erwarteten Beobachtungen. Die Besonderheit gegenüber den in Abschnitt 2.1.1 behandelten merkmalsbasierten Verfahren liegt darin, daß hierbei die Modell-Punktmenge eine kanonische Repräsentation des Objektes sein darf. Das Hauptaugenmerk liegt auf der besonderen Transformation der Punktmengen.

Die Hauptidee des in [Sclaroff & Pentland 95] beschriebenen Verfahrens beruht auf der Beschreibung der Objekte mittels generalisierter Symmetrien, die durch die Eigenwerte zugehöriger finiter Elementmodelle definiert werden. Die resultierende modale Beschreibung kann zur Objekterkennung und -kategorisierung verwendet werden. Dabei wird die Ähnlichkeit zweier Objekte durch den Betrag einer modalen Deformationsenergie ausgedrückt, die zur Überführung der beiden Objekte ineinander benötigt wird. Das Verfahren kann auf Datenmengen beliebiger Dimensionalität angewandt werden. Unter den Testbeispielen finden sich sowohl auf Objektkonturpunkten als auch auf Punkt-

merkmalen basierende Zuordnungs- und Erkennungsexperimente. Als Testobjekte dienen menschliche Hände, Werkzeuge wie Hämmer und Schraubenschlüssel, Flugzeuge sowie Personenkraftwagen vor homogenem Hintergrund. Es ist zu erwähnen, daß die demonstrierten Erkennungsaufgaben zweidimensionaler Natur sind.

In ähnlicher Weise werden auch in [Gold et al. 98] Punktmengen ineinander überführt. Der vorgestellte Algorithmus arbeitet mit deterministischem Abkühlen zur Optimierung geeigneter Gütefunktionen und speziellen Techniken, die im Rahmen der Neuronalen Netze und der statistischen Physik entwickelt wurden. Als Eingabe können verrauschte 2D oder 3D Punktmengen verwendet werden, die von unterschiedlicher Kardinalität sein können und nicht-starren, d.h. verformenden Transformationen unterworfen sein dürfen. Die Testbeispiele zur Erkennung von handgeschriebenen Zeichen demonstrieren die Funktionsweise und Robustheit des vorgestellten Verfahrens. Ähnlich wie bei dem zuvor beschriebenen Ansatz [Sclaroff & Pentland 95] ist damit die Erkennung von Objekten anhand der Konturpunkte bzw. Punktmerkmale leicht denkbar.

2.1.5 Gebietsbasierte Verfahren, Schablonenanpassung

Die in diesem Abschnitt behandelten Verfahren basieren auf der *Anpassung zweidimensionaler Schablonen* an die zugehörigen Bildgebiete oder an die entsprechenden Gebiete geeigneter Merkmalsbilder. Aufgrund der Funktionsweise liegt die Stärke solcher Verfahren ohne weitere Maßnahmen in der Lokalisation zweidimensionaler Objekte bzw. Objektabbilder, vgl. [Sim et al. 99] und [Basri & Jacobs 99].

[Sim et al. 99] verwenden als Merkmalsbilder bzw. für die Schablonen Kantenelementbilder. Als Abstandsmaß zwischen zwei Kantenelementbildern dienen sowohl die Hausdorff-Distanz als auch zwei ihrer Derivate. In einer Computersimulation werden die drei Abstandsmaße anhand synthetischer und realer Daten miteinander verglichen. Als zu lokalisierendes Testobjekt dient ein PKW auf einer Landstraße vor realem, komplexem Szenenhintergrund.

[Basri & Jacobs 99] beschränken sich auf die Erkennung und Lokalisation planarer Objekte unter projektiven Transformationen, wobei partielle Verdeckungen zugelassen sind. Die Modellierung der Objekte erfolgt durch eine geeignete Zerlegung in konvexe Teilgebiete. Die Lokalisationsaufgabe besteht darin, eine Transformation zu finden, welche die konvexen Gebiete des modellierten Objekts auf die zugehörigen Gebiete im Testbild überführt. Es wird gezeigt, daß drei sichtbare Gebiete zur eindeutigen Bestimmung der gesuchten Transformation genügen und für eine große Klasse an Objekten sogar bereits zwei sichtbare Gebiete genügen. Dies ermöglicht insbesondere die robuste Behandlung von Verdeckungen. Mit dem vorgestellten Verfahren lokalisieren die Autoren ein Gemälde von Kandinsky in mehreren Aufnahmen vor einfach gestaltetem Hintergrund.

[Ratan et al. 98], [Ratan et al. 00] verwenden zur Detektion, Erkennung und Lokalisation von Objekten eine zweidimensionale Schablone, die sie geeigneten Abbildungen unterwerfen und mit dem Bild zur Deckung bringen. Die Autoren reduzieren die Lokalisation auf ein eindimensionales Unterproblem in Form einer mittels dynamischer Programmierung realisierten Suche nach der besten Anpassung zwischen Bild und Schablone. Schablone und Bild werden dabei nicht – wie bei der Schablonenanpassung meist üblich – pixelweise subtrahiert. Stattdessen verwenden die Autoren Felder von Intensitätsmustern, die sie nach einer Grauwertnormalisierung durch Grobabtastung und Grobquantisierung gewinnen. Das vorgestellte Verfahren ist so ausgelegt, daß auch bis zu 5 Freiheitsgrade von 3D-Objekten erfaßt werden können. Die Autoren wenden das Verfahren in der Gesichtererkennung an und lokalisieren Personenkraftwagen in Realwelt-Aufnahmen.

Der Artikel [Worthington & Hancock 00] mag in der vorliegenden Literaturvorstellung als Vertreter der gebietsbasierten Verfahren zur Erkennung von 3D-Freiformobjekten anhand derer Schattierungsverteilung dienen. Solche Verfahren sind in der Literatur auch unter dem Kürzel ‘Form aus Schattierung’ (engl.: shape-from-shading) bekannt. Zur Objekterkennung extrahieren die Autoren Regionen uniformer Flächentopographie und verwenden diese zur Erkennung der entsprechenden Objekte in Testbildern. Zur Realisierung werden spezielle Repräsentationen maximaler Flächenstücke verwendet, die bestimmte Unabhängigkeiten gegenüber dem Betrachtungswinkel aufweisen. Der vorgestellte Erkennungsprozeß basiert auf der Betrachtung der Oberflächennormalen. Als Testobjekte dienen z.B. eine Gummiente, eine Tasse oder ein Sparschwein vor homogenem Hintergrund.

2.1.6 Indizierungsbasierte Verfahren

Indizierungs-basierte Erkennungsverfahren arbeiten in der Regel mit speziellen *Strategien zur Durchnummerierung von im Bild extrahierten*, geradlinigen oder nicht geradlinigen *Kanten* dreidimensionaler Objekte. Dies bildet die Grundlage zur Erfassung und Repräsentation der Topologie der Objektkanten bzw. der Objektform. Durch einen Vergleich der aus einem Bild berechneten Daten mit geeigneten, rechnerinternen Objektrepräsentationen kann das Vorhandensein der Objekte geprüft bzw. auf die Lage der Objekte im Raum geschlossen werden. Meist werden bei solchen Verfahren polyedrische Objekte verwendet.

In [Horaud & Sossa 95] findet man einen guten Überblick über die Entwicklungshistorie solcher Verfahren sowie einen klassischen Ansatz zur indizierungsbasierten Objekterkennung unter Verwendung von Streuspeicher-Tabellen zur Beschleunigung. Die Autoren benutzen gewichtete Graphen zur Bildkanten- sowie Objektkodierung und reduzieren die 3D-Erkennungsaufgabe auf graphentheoretische Enthaltenseinsrelationen. Sie machen darauf aufmerksam, daß der Ansatz keine Zuordnungen zwischen Bild- und Objektmerkmalen benötigt und es insbesondere erlaubt, effizient die im Bild enthalte-

nen Objekte mit den entsprechenden Kandidaten aus einer großen Objektdatenbasis identifizieren zu können. Die Ergebnisse werden ausführlich anhand mehrerer polyedrischer Objekte vor homogenem Hintergrund demonstriert.

[Bebis et al. 98] bemängeln an den bekannten indizierungsbasierten Verfahren, daß sie zum Aufbau der verwendeten Streuspeicher-Tabelle entweder eine große Anzahl an Referenzansichten oder 3D-Modelle der Objekte benötigen. Daher schlagen die Autoren einen Ansatz vor, der zum Aufbau der Streuspeicher-Tabelle bestimmte algebraische Funktionen der Objektansichten verwendet. Dadurch kann die Tabelle nach Angaben der Autoren einfach aufgebaut werden, es werden keine Ähnlichkeiten zwischen neuen Objektansichten und den Referenzansichten benötigt, der Ansatz ist zudem allgemeiner gehalten und besser erweiterbar als vorherige Ansätze. Das Verfahren wird eingesetzt zur Lokalisation eines Tesarollers, eines Hefters sowie einer Computermaus ohne erwähnenswerte partielle Verdeckungen vor einem sehr einfachen Hintergrund.

[Beis & Lowe 99] betonen bei dem von ihnen vorgestellten Ansatz, daß das Verfahren nicht – wie bei den meisten Ansätzen in der Literatur – auf der Verwendung bestimmter invarianter Merkmale basiert. Dies erlaubt nach Ansicht der Autoren eine deutlich reichhaltigere Menge an im Erkennungsprozeß berücksichtigten Formen. Zudem wird anstatt der üblichen Streuspeicher-Tabelle eine Baumstruktur verwendet, wodurch der Erkennungsprozeß in hochdimensionalen Räumen (wie sie oft bei großen Modelldatenbasen verwendet werden) besonders effizient gestaltet werden kann. Die Effizienz des Verfahrens wird dabei hauptsächlich durch eine modifizierte Baumsuche realisiert, die vergleichbar ist mit einer Suche nach dem nächsten Nachbarn. Solche Nachbarn werden zu Wahrscheinlichkeitsschätzungen herangezogen, die eine geeignete Sortierung der zu verifizierenden Lagehypothesen und so eine beträchtliche Aufwandsreduzierung erlauben. Die Autoren verwenden als Lokalisationsbeispiele einen Hefter, eine Art Schachtel, und eine Wäscheklammer und arbeiten mit partiellen Verdeckungen sowie komplexen Hintergründen.

Während die oben behandelten Ansätze auf geradlinigen Objektkanten basieren, wenden [Chen & Stockman 98] die Indizierung sogar zur modellbasierten Erkennung und Lokalisation von dreidimensionalen *Freiform*-Objekten an. Dazu untergliedern die Autoren die Erkennungsaufgabe in drei Teile: ein Modellierungsteil, ein Anpassungsteil und die eigentliche Indizierung. Der angesprochene Modellierungsteil generiert zu verschiedenen Betrachtungswinkeln die für die Erkennungsaufgabe notwendigen Aspekte der Objektkontur, vgl. Abschnitt 2.1.8. Der Anpassungsteil richtet die zu einer initial vorgegebenen, ungefähren Objektlage gehörige Modellkontur nach den im Bild beobachteten Kantenstrukturen aus, indem eine geeignete Bewertungsfunktion iterativ minimiert wird. Der zentrale Teil der Indizierung generiert Hypothesen, die sowohl den Objektkandidaten als auch die im Anpassungsteil benötigte ungefähre Objektlage sowie die Skalierung beinhalten. Zur Beschleunigung des Verfahrens werden die Hypothesen anhand von Auftrittswahrscheinlichkeiten sortiert, die aus a-priori Kenntnissen gewonnen worden sind. Um das Indizierungsverfahren auf die auftretenden Freiformkonturen anwenden zu können, werden die Konturen zuvor geeignet in Einzelteile zerlegt. Die

Autoren verwenden 20 dreidimensionale und 80 zweidimensionale Testobjekte, die sie in einer insgesamt 658 Modellaspekte umfassenden Datenbasis repräsentieren. Die abgedruckten Testbeispiele, wie z.B. eine nicht polyedrische Fahrzeugkarosserie, verschiedene Tiere sowie ein Telefonhörer, überzeugen zum einen durch den Realitätsbezug der Formen. Zum anderen weisen einige Beispiele das Verfahren als robust gegenüber z.T. signifikanten partiellen Verdeckungen aus; statistische Untersuchungen wurden beigefügt. Die Autoren sprechen sogar von einer hohen Unempfindlichkeit des Verfahrens gegenüber einem komplexen Hintergrund, was allerdings durch keine Abbildungen belegt wird.

2.1.7 Verfahren mittels Hough-Transformation

Einigen Erkennungs- und Lokalisationsverfahren in der Fachliteratur liegt die *Hough-Transformation* oder eine ihrer Generalisierungen zugrunde. Ein im Hinblick auf die zu lösende Aufgabe herausragender Vorteil dieser Transformationen besteht darin, daß geometrische Primitive im zu bearbeitenden Eingabebild im Ergebnisraum, dem sogenannten Parameterraum, auf einfache, charakteristische Strukturen, wie beispielsweise einzelne Punkte fallen. Diese Strukturen können besonders gut für die zu erledigende Erkennungs- oder Lokalisationsaufgabe verwendet werden. Zur genaueren Funktionsweise der Hough-Transformation sei auf [Kassim et al. 99] verwiesen. Dort findet sich eine empfehlenswerte Zusammenstellung einiger gängiger generalisierter Hough-Transformationen samt einem ausführlichen Vergleich. Da die Berechnung der Hough-Transformation insbesondere bei komplexen Anwendungen sehr zeitaufwendig sein kann, geben [Kiryati et al. 00] zwei Möglichkeiten zur Beschleunigung an. Die eine kombiniert die Hough-Transformation mit einer zufallsgesteuerten Datenabtastung, und die andere verwendet einen probabilistischen Ansatz. Aufgrund der Funktionsweise der Hough-Transformation liegt die Stärke Hough-basierter Verfahren in der Regel in der 2D-Erkennung bzw. -lokalisierung, siehe unten. Durch entsprechende Erweiterungen sind jedoch auch 3D-Lokalisierungen realisierbar, wie z.B. [Silberberg et al. 84] zeigt. Allerdings muß hier vorausgesetzt werden, daß die Objektgröße im Bild bekannt ist. Mittels iterativer Anwendung des vorgestellten Hough-Verfahrens können die verbleibenden 5 Parameter zur Schätzung der Objektlage bestimmt werden.

[Guil et al. 99] stellen einen als typisch zu bezeichnenden Ansatz zur Hough-basierten Detektion zweidimensionaler Objekte vor. Die generalisierte Hough-Transformation wird verwendet, um eine vorgegebene Schablone im Bild wiederzufinden. Das vorgestellte Lokalisationsverfahren berechnet dabei sowohl Position und Orientierung der Schablone im Bild als auch deren Skalierung. Als Testobjekte werden ein Dreieck, ein Strichmännchen sowie eine Spitzzange vor nicht-trivialem Hintergrund verwendet.

[Chau & Siu 99] entwickeln eine neue Variante einer generalisierten Hough-Transformation zur Verbesserung von Detektionsgenauigkeit und Rechenzeit. Die Autoren berechnen einen sog. charakteristischen Winkel, den sie zur Generierung der Hough-Tabelle

verwenden, um Beschaffenheit und Zugriffseigenschaften der Tabelle zu optimieren. Als Testbeispiele werden zweidimensionale, sich zum Teil hochgradig überdeckende Objekte verwendet, deren Ränder hauptsächlich aus linien- oder kreisförmigen Segmenten bestehen.

Um neben zweidimensionalen Objekten auch perspektivisch transformierte 2D-Objekte erkennen und lokalisieren zu können, erweitern [Lo & Tsai 97] die generalisierte Hough-Transformation so, daß sie invariant gegenüber perspektivischen Transformationen der Objekte wird. Die vorgestellte Hough-Transformation basiert auf einer perspektivischen Referenztafel. Diese wird aufgebaut, indem auf die vorgegebene Schablone zu allen Beobachtungspositionen der Kamera perspektivische Transformationen und dazu inverse Abbildungen angewandt werden. Das Verfahren weist damit eine deutliche Parallele auf zu den betrachtungswinkelbasierten Verfahren aus Abschnitt 2.1.8. Kandidaten für das zu erkennende Objekt ergeben sich durch Schwellwertvergleich mit den Spitzen in der Hough-Tabelle (bzgl. der perspektivischen Abbildungen). Der Kandidat, der nach Anwendung der entsprechenden inversen perspektivischen Transformation zu der besten Anpassung zwischen Modell und Bildobjekt führt, liefert schließlich das Ergebnis samt der gesuchten perspektivischen Transformation. Als Testbeispiele werden von den Autoren nicht verdeckte Spielzeugflugzeuge, Schlüssel und beschriftete Telefon- oder Kreditkarten vor homogenem Hintergrund verwendet.

[Olson 99] unterscheidet zwischen zwei Klassen von Verfahren zur Modellanpassung und -extraktion (zu denen bekanntlich auch die Verfahren zur Detektion, Erkennung und Lokalisation von Objekten gehören). Die eine Klasse wird von den hier behandelten Hough-basierten Verfahren gebildet. Die zweite Klasse faßt Olson unter dem Begriff Generiere-und-Teste-Verfahren zusammen. Er bemerkt, daß beiden Klassen prinzipielle Probleme inhärent sind: Während die Generiere-und-Teste-Verfahren besonders Empfindlichkeiten gegenüber Störungen der Eingabedaten aufweisen, sind die Hauptprobleme der Hough-basierten Verfahren bei komplexen Anwendungen die hohen Rechenzeit- und Speicheranforderungen sowie die Häufigkeit falscher Alarme. Olson beschreibt daher ein allgemeines Verfahren zur Kombination beider Verfahrensweisen, die nach seinen Angaben beiden Ansätzen überlegen ist. Er gibt abschließend ein Verfahren zur Objekterkennung an und wendet dieses auf das Bild eines Hefters in einer Realwelt-Szene (Hintergrund nicht komplex) an.

2.1.8 Betrachtungswinkelbasierte Verfahren

Wie der Name schon sagt, beruht die *betrachtungswinkelbasierte Objekterkennung* (engl.: view based recognition) auf dem *vom Betrachtungswinkel abhängigen Erscheinungsbild* der zu erkennenden Objekte. Oftmals verlangen solche Verfahren daher bei der Erkennung und Lokalisation dreidimensionaler Objekte die Verwendung von Datenbasen, welche für alle Gitterpunkte einer geeigneten Rasterung des Raumes aller zu erkennenden Orientierungen bzw. Positionen eines betreffenden Objektes einen entsprechenden Repräsentanten enthalten. [Vel & Aeberhard 00] geben einen interessanten

Literaturüberblick über die unterschiedlichen Stoßrichtungen betrachtungswinkelbasierter Verfahren.

[Borotschnig et al. 00] arbeiten bei zwei Freiheitsgraden je Objekt mit 540 unterschiedlichen Betrachtungspositionen zu je 40 aufgenommenen, leicht unterschiedlichen Objektrepräsentanten. Der Ansatz verwendet als betrachtungswinkelbasierte Objektrepräsentation den sogenannten parametrischen Eigenraum (siehe [Murase & Nayar 95], [Borotschnig et al. 00]) und erweitert diesen durch die Verwendung von Wahrscheinlichkeitsverteilungen. Dies ermöglicht es, die den Eingabebildern inhärenten Variationen verschiedener Art einzufangen. Eine über die reine Erkennungsaufgabe hinausgehende Objektlokalisierung wird bewerkstelligt durch die Generierung künstlicher Lernbeispiele mittels Interpolation im Eigenraum. Als Testbeispiele verwenden die Autoren Spielzeugautos und -tiere vor homogenem Hintergrund.

[Winkeler et al. 99] beschränken sich auf einen rotatorischen Freiheitsgrad und verwenden bei den Testbeispielen 120 Betrachtungswinkel je Objekt, d.h. eine Rasterung des Orientierungsraumes von 3° . Das vorgestellte System wählt während der Konstruktion der Eigenraumbasis unter den Objektansichten automatisch eine repräsentative Untermenge aus. Zur Objektlokalisierung muß die Kamera systemgesteuert bewegt werden. Als Testbeispiele dienen nicht-triviale Objekte (z.B. Flasche, Ventilator, Glühbirne, Basketball) vor homogenem Hintergrund, wobei unterschiedliche Beleuchtungssituationen berücksichtigt werden.

[Huttenlocher et al. 99] arbeiten ebenfalls mit dem parametrischen Eigenraum. Die Autoren bemerken, daß die bekannten betrachtungswinkelbasierten Verfahren zwar bei einer Reihe von Erkennungsaufgaben erfolgreich eingesetzt werden können, daß aber die Verfahren deutliche Schwächen zeigen, wenn die zu erkennenden Objekte partiell verdeckt sind oder vor einem komplexen Hintergrund erscheinen. Sie entwickeln ein gegenüber diesen Schwächen tolerantes Verfahren, indem sie die Eigenraumtechnik mit der generalisierten Hausdorff-Distanz [Huttenlocher et al. 93] kombinieren. Das Verfahren wird angewandt auf Merkmalsbilder, im Beispiel auf ein Kantenbild. Als Testbeispiel dient eine partiell verdeckte, beschriftete Medikamentenschachtel in einer Realweltszene.

[Hall et al 00] wenden sich ebenfalls der Lösung der Schwierigkeiten bei partiell verdeckten Objekten vor komplexem Hintergrund zu. Das vorgestellte Verfahren basiert auf der Abbildung *lokaler* Gegebenheiten des betrachtungswinkelabhängigen Erscheinungsbildes auf funktionale Flächen in einem geeigneten Deskriptorraum. Ein Objekt wird so für jeden Betrachtungswinkel durch eine Menge solcher Flächen repräsentiert. Der Erkennungsprozeß ist dann äquivalent zu einer Suche nach korrespondierenden Deskriptorflächen in einer Datenbasis, deren Flächen zur Beschleunigung strukturell geordnet wurden. Die gefundene Fläche enthält mitgespeicherte Informationen über das Objekt sowie dessen Position und Orientierung und liefert so direkt alle zur Erkennung und Lokalisation notwendigen Daten. Das Verfahren operiert auf Farbbildern und wird eingesetzt für viele verschiedene, sich teilweise überdeckende Schreibtischu-

tensilien (Papier, Hefter, Tesaroller, Kugelschreiber, Geodreieck, Spitzer, Radiergummi, Schere) auf einem Schreibtisch im Hinblick auf das Ziel, solche Objekte und deren Zusammenspiel zur Mensch-Maschine-Interaktion benutzen zu können.

[Vel & Aeberhard 00] erweitern ein betrachtungswinkelbasiertes Erkennungsverfahren, das auf der Klassifikation von geeignet im zweidimensionalen Objektabbild platzierten Linien basiert. Die Autoren sprechen von hohen Generalisierungsraten bei der Objekterkennung und können durch diese Eigenschaft die bei betrachtungswinkelbasierten Erkennungsverfahren zumeist hohen Trainings- und Erkennungszeiten reduzieren. Als Testobjekte fungieren u.a. ein Bügeleisen, ein Tastentelefon oder eine Teekanne vor homogenem Hintergrund.

Eine Brücke zwischen betrachtungswinkelbasierten und auf Neuronalen Netzen basierenden Verfahren (vgl. Abschnitt 2.1.10) schlagen [Yang et al. 00], indem sie einen betrachtungswinkelbasierten Lernalgorithmus für ein aus linearen Einheiten aufgebautes, spärliches Neuronales Netz angeben, das in der Eingabeschicht einfache Relationen realisiert. Als Eingabe des Neuronalen Netzes dienen grauwertbasierte und grauwertkantenbasierte Repräsentationen. Als Testobjekte dienen viele verschiedene Objekte, wie z.B. Spielzeugautos, Spielzeugtiere, Gefäße, Schachteln usw. vor homogenem Hintergrund.

2.1.9 Probabilistische Verfahren

Statistische und wahrscheinlichkeitstheoretische Grundlagen, insbesondere die *Bayes'sche Regel*² bilden die Basis der im Folgenden behandelten Verfahren. Die Bayes'sche Regel stellt dabei die grundlegende Beziehung her zwischen gesuchten Wahrscheinlichkeiten und einfacher bestimmbar bzw. zu lernenden Verteilungen.

In [Sullivan et al. 00] bildet ein Ansatz der Schablonenanpassung die Grundlage zur Objektlokalisierung: Die Lokalisierung eines Objektes wird mit der Suche nach einer Transformation gleichgesetzt, die eine vorgegebene Schablone in das Objektabbild in der untersuchten Aufnahme überführt. Die zulässigen Transformationen dürfen dabei nicht nur affine Abbildungen sondern auch diverse Verzerrungen umfassen. Zur Berechnung der passenden Transformation modellieren die Autoren die Transformationen wahrscheinlichkeitstheoretisch und berechnen mit Hilfe der Bayes'schen Regel die zugehörige a-posteriori Verteilung aus entsprechenden a-priori Wahrscheinlichkeiten. Das Hauptproblem des Verfahrens besteht folglich in der Bestimmung geeigneter statistischer Modelle für das a-priori Wissen bezüglich Vorder- und Hintergrund. Die Autoren konzentrieren sich in der Arbeit vor allem auf das wesentlich schwierigere Vordergrundmodell, wobei auch Beleuchtungsvariationen berücksichtigt werden. Das vorgestellte Verfahren wird angewandt zur Lokalisierung und Verfolgung von Gesichtern oder einer Hand vor hochgradig komplexem Hintergrund. Ferner werden verschiedene Objekte,

² $p(A|x) = \frac{p(x|A) \cdot p(A)}{p(x)}$.

wie beispielsweise eine Vase oder eine Geige in Realwelt-Szenarien erfolgreich lokalisiert. Die angegebenen Beispiele beeindrucken vor allem aufgrund der Komplexität des jeweils verwendeten, zum Teil sogar texturierten Hintergrundes. Zudem sind bei einigen Testbeispielen auf den zu lokalisierenden Objekten starke Schatten zu beobachten, was auf eine deutliche Robustheit des Verfahrens gegenüber Beleuchtungsvariationen hindeutet. Eine Tauglichkeit des Verfahrens zur Lokalisation frei im Raum bewegter dreidimensionaler Objekte kann den Testbeispielen allerdings nicht entnommen werden. Die Stärke des Verfahrens scheint daher in der Lokalisation zweidimensionaler Objekte bzw. Objektansichten zu liegen. In einer ähnlichen, am gleichen Institut entstandenen und ebenfalls sehr interessanten Arbeit [Sullivan et al. 99] wenden sich die Autoren sogar starken Verdeckungen einer zu lokalisierenden Hand vor hochgradig komplexem Hintergrund sowie der Fußgängerverfolgung zu. Besonders eindrucksvoll ist die Lokalisation und Verfolgung eines energisch vom Wind bewegten Blattes eines Baumes inmitten weiterer Blätter.

[Olson 98] bemerkt, daß die Hausdorff-Distanz bei Anwendungen der Bildauswertung zwar eine beliebte und mächtige Basis darstellt, daß jedoch in diesen Anwendungen keine probabilistischen Modelle zum Einsatz kommen. Dadurch sei insbesondere die formale Behandlung von a-priori Wissen sowie die Berücksichtigung von Unsicherheiten in den verwendeten Merkmalen drastisch eingeschränkt. Zur Beseitigung dieser Nachteile stellt Olson eine probabilistische Formulierung des Anpassungsproblems vor, welche mit der Schätzung der maximalen Mutmaßlichkeit arbeitet und eine Generalisierung der Hausdorff-basierten Anpassung darstellt. Darauf aufbauend gelingt es dem Autor, eine effiziente Objektlokalisierung zu realisieren, indem er eine hierarchische Suchstrategie auf den (diskretisierten) Raum möglicher Modellpositionen anwendet [Olson 98], [Olson 00]. Bei dieser Suche werden sukzessive Teilräume des Suchraumes in geeigneter Weise bewertet und gemäß ihrer Bewertung entweder verworfen, oder sie werden unterteilt und die entstandenen Teilräume werden anschließend weiter untersucht. Olson wählt bewußt *kein iteratives*, lokales *Suchverfahren*. Er weist darauf hin, daß die angesprochene Suchstrategie eine *globale Suche* durchführt und als Ergebnis die tatsächlich *beste Modellposition* findet. Das entwickelte Verfahren wird auf die Selbstlokalisierung (d.h. Standortbestimmung) eines mobilen Roboters im Gelände des Planeten Mars angewandt.

[Wheeler & Ikeuchi 95] nutzen a-priori Wissen, um die Anzahl an zu testenden Lagehypothesen bei einer merkmalsbasierten Objekterkennung bzw. -lokalisierung zu verringern (vgl. Abschnitt 2.1.1). Dazu schränken die Autoren mittels statistischer Methoden unter Verwendung von Markovfeldern (engl. Markov Random Fields) die hypothetisch möglichen Zuordnungen zwischen Modell- und Bildmerkmalen ein. Das vorgestellte Lokalisationsverfahren funktioniert nach Angabe der Autoren trotz partieller Verdeckung der Objekte oder eines signifikanten Fehlers in der initialen Lageschätzung zuverlässig. Als Testbeispiele werden relativ einfache, polyedrische Teile sowie ein Hefter und ein Tesaroller vor mehr oder weniger komplexem Hintergrund verwendet.

[Boykov & Huttenlocher 99] wenden sich ebenfalls der merkmalsbasierten Objekter-

kennung unter Verwendung von Markovfeldern und MAP-Schätzungen (Maximum Aposteriori Schätzungen) zu. Boykov und Huttenlocher geben effiziente Lösungen an für ein breites Spektrum an a-priori Wissen, das die Abhängigkeiten zwischen einzelnen Objektmerkmalen explizit modelliert. Dadurch können insbesondere aufgrund partieller Überdeckungen nicht zugeordnete Merkmale adäquat behandelt werden. Als Schwerpunkt stellen die Autoren ein approximatives Anpassungsverfahren vor, das die örtliche Korrelation benachbarter Objektmerkmale berücksichtigt. Abschließend belegen die Autoren die Zuverlässigkeit ihres mit der Hausdorff-Distanz kombinierten Verfahrens bei hochgradig komplexem Hintergrund und partiellen Verdeckungen bis zu 40%. Leider wird als Testobjekt lediglich eine sehr einfache, geschlossene Strichzeichnung verwendet.

Ein völlig anderer Weg wird in [Herbin 98] besprochen. Der Autor präsentiert eine generalisierte, probabilistische Version eines Konzeptes, welches auf Aspektgraphen basiert. Solche Ansichtsgraphen werden üblicherweise verwendet, um die Variation einer Objektansicht während einer Änderung der Beobachtungsposition oder einer Änderung der Sensorparameter zu repräsentieren. Die Erkennung dreidimensionaler Objekte wird in Operationen auf Markov-Ketten überführt. Das Verfahren wird angewandt zur 3D-Erkennung ziviler und militärischer Flugzeuge vor homogenem Hintergrund. Die theoretisch erwartete sowie die in der Praxis tatsächlich erzielte Erfolgsrate werden ausführlich behandelt und mit entsprechenden Grafiken illustriert.

2.1.10 Neuronale Netze verwendende Verfahren

Die in diesem Abschnitt zitierten Arbeiten demonstrieren, wie unterschiedlich *Neuronale Netze* zur Lösung der Erkennungs- bzw. Lokalisationsaufgabe eingesetzt werden können.

[Toth & Culverhouse 99] verwenden bestimmte Grobkodierungen von Merkmalsinformationen auf mehreren Auflösungsstufen. Klassische Methoden der Merkmalsextraktion werden verbunden mit den Vorzügen der Wavelet-Transformation und die Vorteile der mit mehreren Kanälen und Auflösungen arbeitenden Architektur erklärt. Zur Objektkategorisierung werden die extrahierten Merkmale einem mehrschichtigen Neuronalen Netz eingespeist, dessen Parameter mit Fehlerrückpropagierung bestimmt werden. Das System wird anhand von 2D-Ansichten verschiedener Prismen und Kegel unterschiedlicher Grundflächentopologie vor homogenem Hintergrund ohne Berücksichtigung partieller Verdeckungen getestet. Außerdem wird das Verfahren zur Erkennung dreier Fischlarvenarten vor homogenem Hintergrund eingesetzt, d.h. zur Erkennung nicht-starrer Objekte.

[Yáñez-Suárez & Azimi-Sadjadi 99] schlagen mit dem vorgestellten Verfahren eine Brücke zwischen den Hough-basierten Erkennungsverfahren aus Abschnitt 2.1.7 und den auf Neuronalen Netzen basierenden, indem sie die Ausgabe der Hough-Transfor-

mation als Eingabe eines Neuronalen Netzes verwenden. Zunächst werden dazu die Konturen der Objektabbilder durch Polygone approximiert und in den Hough-Raum transformiert. Anschließend generiert ein spezielles Neuronales Netz Ballungen kollinear oder paralleler Kanten, welche abschließend als Basis zur Identifikation der evtl. partiell verdeckten Objekte dienen. Im aufgeführten Anwendungsbeispiel werden Stäbe in komplexen Szenen lokalisiert.

2.1.11 Sonstige Verfahren

[Lee et al. 99] stellen einen Ansatz zur Erkennung dreidimensionaler Objekte vor, der auf der Verwendung von Superquadriken basiert. Dazu werden die zu erkennenden Teile mittels einer Menge volumetrischer Primitive in Form von Superquadriken und deren Zusammenhangsstruktur beschrieben. Außerdem werden in der Modelldatenbasis Oberflächeninformationen wie Flächentyp sowie die Zusammenhangsstruktur der Oberflächenteile mit abgelegt. Der vorgestellte Erkennungsalgorithmus wird erfolgreich zur Erkennung und Lokalisation von Werkstücken beliebiger Lage eingesetzt. Die abgebildeten Werkstücke sind meist aus zwei bis drei quader-, zylinder- oder kugelförmigen Teilen zusammengesetzt.

Auf der Analyse und Interpretation von Strichzeichnungen basiert ein Ansatz von Wang [Wang 99], der auf aus Grauwertbildern extrahierte Kantenelementbilder angewandt werden kann. Der vorgestellte strukturelle Ansatz verwendet Methoden zur Darstellung von Objektansichten mittels geeigneter Linearkombinationen aus zuvor gespeicherten Ansichten. Objekte werden identifiziert, indem spezielle Schablonen aus zweidimensionalen Bildmodellen konstruiert werden. Das Verfahren eignet sich zur Repräsentation, Visualisierung sowie Interpretation von Strichzeichnungen als 3D-Objekte und insbesondere zur Erkennung und Lokalisation von dreidimensionalen, mehrgliedrigen, aus polyedrischen Teilen bestehenden Objekten. Als Testbeispiele dienen u.a. verschiedene Schachteln mit unterschiedlichen Öffnungswinkeln der Deckel, ein aus mehreren Polyedern und einem Zylinder zusammengesetzter, mehrgliedriger Panzer sowie eine Schere unter verschiedenen Öffnungswinkeln. Störkanten durch komplexe Hintergrundstrukturen, partielle Verdeckungen o.ä. werden nicht behandelt.

In [Batlle et al. 00] geht es um die Zusammenstellung und Klassifikation diverser Verfahren aus der Literatur zur Objekterkennung und Segmentierung in Farbbildern, die im Freien aufgenommen wurden. Die Autoren unterscheiden zwischen ‘bottom-up’-, ‘top-down’- und hybriden Ansätzen. Die angegebenen Beispiele basieren größtenteils auf Segmentierungsansätzen und beziehen sich auf die Erkennung von Objekten in Landschaftsbildern bewohnter oder nicht bewohnter Gegenden bzw. in Straßenszenen. Mit diesem Artikel soll die Sparte farbbasierter Verfahren angesprochen werden, die hier ansonsten nicht näher erläutert werden soll.

Es sei erwähnt, daß theoretisch natürlich auch Verfahren zur *externen Kamerakalibrie-*

zung zur Lokalisation von Objekten eingesetzt werden können, schließlich spielt es keine Rolle, ob die Lage des Objektes zur Kamera oder die Lage der Kamera bezüglich des Objektes berechnet wird. Dazu müssen die verwendeten Merkmale alle vom zu lokalisierenden Objekt stammen und den Voraussetzungen der Kalibrieralgorithmen genügen. Hauptschwierigkeiten sind dabei natürlich die Extraktion der richtigen Merkmale aus dem Bild sowie die Berücksichtigung komplexer Szenen und partieller Verdeckungen.

2.1.12 Zusammenfassung und Diskussion

Die in den vorhergehenden Abschnitten angesprochenen Artikel zeigen, wie breitgefächert die Ansätze zur Objekterkennung und -lokalisierung sind. Die angeführten Literaturzitate stellen eine Vielzahl unterschiedlicher Verfahrensweisen mit unterschiedlichen Schwerpunkten vor und lösen viele unterschiedliche Teilprobleme. Viele der vorgestellten Verfahren wurden nicht bei komplexen Realwelt-Hintergründen oder partiellen Verdeckungen getestet und weisen hier Schwachstellen auf. Abgesehen davon können aus der prinzipiellen Funktionsweise folgende Verfahrens-inhärente Stärken bzw. Schwächen abgeleitet werden:

Merkmalsbasierte und indizierungsbasierte Verfahren: Problematisch bleiben fehlerhafte oder mehrdeutige Zuordnungen in komplexen Szenen. Ferner besteht eine Abhängigkeit der Lokalisation von der Güte der Bildmerkmale sowie insbesondere der Modellmerkmale.

Freiformkurven- bzw. freiformflächenbasierte Verfahren: Die Stärke liegt natürlich in der Erkennung bzw. Lokalisation frei geformter 3D-Objekte in einer Szene.

Überführung von Punktmengen ineinander: Ein großer Vorteil besteht in der Verwendbarkeit *generischer* Modelle zur Lokalisation unterschiedlich ausgeprägter Vertreter definierter Objektklassen. Besonders schwerwiegende Probleme stellen sich allerdings durch komplexe Realwelt-Hintergründe und Verdeckungen.

Gebietsbasierte Verfahren, Schablonenanpassung: Diese eignen sich besonders gut zur Lokalisation zweidimensionaler (evtl. komplexer) Objekte oder Objektansichten. Eine robuste 3D-Lokalisation von Objekten in komplexen Szenen bleibt dagegen problematisch.

Verfahren mittels Hough-Transformation: Hough-basierte Verfahren eignen sich vorwiegend zur Lokalisation von 2D-Objekten und sind in der Regel unempfindlicher gegenüber Rauscheinflüssen, schlecht ausgeprägten Bildmerkmalen und Ausreißern als viele andere Verfahren. Als Nachteil ist allerdings zu erwähnen, daß einige Hough-basierte Verfahren bei komplexen Anwendungen hohe Rechenzeit- und Speicheranforderungen aufweisen. Ferner ist die Häufigkeit falscher Alarme ein prinzipielles Problem, besonders bei der Erkennung und Lokalisation von 3D-Objekten in komplexen Realweltszenen.

Betrachtungswinkelbasierte Verfahren: Die Stärke gegenüber den meisten anderen Ansätzen liegt darin, daß insbesondere auch komplexe Objekte (bei denen z.B. Kantenrepräsentationen schwierig sind) behandelt werden können. Möglicherweise kann allerdings die Datenbasis schwer alle möglichen Variationen wie Beleuchtungsänderungen, Schattenwurf u.a. in Realweltbildern einfangen.

Probabilistische Verfahren: Als wohl wichtigster Vorteil ist die effiziente Lokalisation zu nennen, die selbst unter Realwelt-Bedingungen in den meisten Fällen funktioniert. Allerdings ist damit zu rechnen, daß bzgl. des verwendeten a-priori Wissens ‘ungewöhnliche’, aber dennoch zutreffende Objektlagen existieren, die nicht erkannt werden.

Neuronale Netze verwendende Verfahren: Das prinzipielle Problem bei Neuronalen Netzen ist, daß man nicht weiß, ob tatsächlich das Gewünschte gelernt wurde. Im Fehlerfall kann man das Problem nur schwer analysieren und selten lokal beheben.

Es ist zu bemerken, daß die Literatur nur äußerst spärlich Verfahren anbietet, die mit *Abtastverfahren im Szenenraum* arbeiten, obwohl solche Verfahren relativ nahelegend erscheinen. Dies dürfte vor allem auf Effizienzaspekte zurückzuführen sein, da die Rechenzeit exponentiell von der Größe des Suchraumes bzw. der Abtastgenauigkeit abhängt. An dieser Stelle sei auf die Literaturstellen [Olson 98] bzw. [Olson 00] aus Abschnitt 2.1.9 verwiesen, bei welchen der Autor bewußt eine hierarchisch erschöpfende Suche im Szenenraum einsetzt. Einen ähnlichen Ansatz verwendet Jurie in [Jurie 98a] und [Jurie 99] in kleinen Parameterräumen zur Realisierung eines robusten, lagekorrigierenden Verifikationsalgorithmus, vgl. Abschnitt 2.1.1.

2.2 Personenverfolgung

[Kakadiaris & Metaxas 00] geben eine sehr ausführliche Zusammenfassung der verschiedenen Verfahren zur Personenverfolgung aus der Literatur bis 1999 an. Da dem kaum etwas hinzuzufügen ist, sei für Verfahren bis 1999 auf die dortige Literaturübersicht verwiesen. Lediglich folgende Arbeiten sind an dieser Stelle zu ergänzen.

[Kakadiaris & Metaxas 00] stellen einen modellbasierten Ansatz zur 3D-Verfolgung von Personen vor, der rein Bildkontur-basiert arbeitet. Zur Vermeidung von Mehrdeutigkeiten bei der Lageschätzung, wie sie bei Verwendung von nur einer Kamera entstehen, verwenden die Autoren eine trinokulare Kameraanordnung mit drei paarweise orthogonalen Betrachtungsrichtungen. Als Personenmodell wird eine deformierbare Repräsentation gewählt und während der Anpassung so verändert, daß sich eine möglichst gute Anpassung an die Silhouette der Person in den drei Aufnahmen ergibt. Mögliche Verdeckungen der Körperteile während der Bewegung werden explizit berücksichtigt. Das

Verfahren wird zur Verfolgung des Oberkörpers, des Kopfes sowie der Arme einer Person eingesetzt. Laut den Autoren steht aus theoretischer Sicht einer Verfolgung des gesamten Körpers nichts im Weg.

Zur Erkennung von Handzeichen und Gesten stellen [Huang et al. 00] ein geeignetes Modell vor zur Beschreibung der auftretenden Formen sowie deren Veränderlichkeit über die Zeit hinweg. Das Modell wird aufgebaut mit Hilfe eines Algorithmus zum Lernen geeigneter Muster aus Trainingsbildfolgen. Dadurch können in einer Testbildfolge nur Bewegungen erkannt werden, zu denen eine ähnliche, entsprechende Repräsentation in der Modellierung existiert. Übergänge zwischen den repräsentierten Formen werden mittels eines verdeckten Markovmodells beschrieben. Das Verfahren wird eingesetzt zur Erkennung zahlreicher Handzeichen unter Verwendung komplexer Hintergründe.

[Sidenbladh et al. 00a], [Sidenbladh et al. 00b] verwenden aus Kegelstümpfen und Kugeln zusammengesetzte 3D-Personenmodelle und projizieren die Silhouette zur (monokularen) Verfolgung in die Bildebene. Die Anpassung erfolgt hierbei mit Hilfe probabilistischer Methoden zur Berücksichtigung auftretender Effekte im Grauwertbild sowie zur Bewegungsmodellierung. Eine Schätzung des Optischen Flusses wird einbezogen. Das Verfahren wird zur Verfolgung von gehenden Personen in Realweltszenen sowie zur Verfolgung von Armbewegungen eingesetzt.

Durch Vergleich der hier bzw. in [Kakadiaris & Metaxas 00] zitierten Ansätze mit [Wachter 97] findet man schnell einige Parallelen zwischen Personenmodellierung, Funktionsweise sowie den erhaltenen Resultaten. Dies spricht dafür, daß die in der vorliegenden Dissertation durchgeführten Untersuchungen zur Personenverfolgung dem aktuellen Stand der Forschung entsprechen.

2.3 Verfolgung/Regelung

Da in [Tonko 97] die wichtigsten Artikel vor 1998 bereits behandelt wurden, sei bezüglich Arbeiten dieser Zeit auf die dortige Literaturbearbeitung verwiesen. Zur Behandlung der Artikel seit 1998 dienen die gleichen Literaturquellen wie bei der Recherche zur Erkennung und Lokalisation bzw. wie bei der Personenverfolgung (Rosenfeld, Bibliotheken, Internet).

[Drummond & Cipolla 00] stellen ein echtzeitfähiges Verfahren zur Verfolgung mehrerer oder mehrgliedriger Objekte in Bildfolgen vor. Dabei können eine oder mehrere Kameras verwendet werden. Zur Berücksichtigung der dabei auftretenden Beschränkungen wird eine allgemeine Methode vorgestellt, welche die geometrischen Zusammenhänge zwischen Objekten beschreibt, z.B. wenn Objekte über Scharniere zusammenhängen oder über eine gemeinsame Grundfläche miteinander verbunden sind. Mit dem vorgestellten Ansatz wird ein Objekt verfolgt, das aus drei miteinander über Scharniere

verbundenen, quadratischen Platten besteht. Zudem wird das in Abschnitt 4.4.3 behandelte Schiffsteil verfolgt.

[Marchand et al. 99] beschreiben einen echtzeitfähigen Ansatz zur Verfolgung polyedrisch modellierter Objekte. Die Lageschätzung geschieht durch Anpassung des ins Bild projizierten Modells an das Gradientenbetragsbild, indem eine nicht-lineare, von drei Lageparametern abhängende Gütefunktion iterativ minimiert wird. Das Verfahren wird zur Verfolgung eines Quaders sowie einer Schraubenmutter in Realweltszenen vor z.T. komplexem Hintergrund eingesetzt.

Das in [Black & Jepson 98] vorgestellte Verfahren eignet sich zur Verfolgung starrer wie auch mehrgliedriger Objekte und basiert auf einer betrachtungswinkelabhängigen Repräsentation der zu verfolgenden Objekte (vgl. Abschnitt 2.1.8). Um das zu erreichen, wird die Eigenraum-Technik sowie ein Verfahren zur Schätzung des Optischen Flusses geeignet erweitert. Die Eigenraumtheorie wird im Sinne der Robustheit geeignet modifiziert. Bei der Schätzung des Optischen Flusses muß sowohl die Objektansicht als auch die affine Transformation zwischen Eigenraum und Bild berücksichtigt werden. Das Verfahren wird zur Verfolgung einer von einem Menschen im Raum bewegten Coladose sowie zur Erkennung von Handzeichen eingesetzt.

[Hsu & Aquino 99] entwickeln ein Verfahren zur robusten, monokularen Regelung bei stationärer Kamera, das mit ungenauen Roboter- und Kameraparametern auskommt. Dazu wird ein adaptives Verfolgungsverfahren vorgestellt sowie ein adaptives Regelungssystem, welches die nichtlineare Dynamik der Anordnung berücksichtigt.

Mit ähnlicher Zielsetzung wie [Hsu & Aquino 99] realisieren [Zergeroglu et al. 99a], [Zergeroglu et al. 99b] eine robuste, sichtsystemgestützte Regelung, welche den Einfluß von Unsicherheiten aufgrund der Roboterdynamik oder einer ungenauen Kamerakalibrierung kompensiert. Das Verfahren kann sowohl bei Systemen mit stationärer Kamera eingesetzt werden als auch bei Kamera-in-Hand-Konfigurationen.

[Benameur & Bélanger 98] setzen eine sichtsystemgestützte Regelung ein, um ein sich bewegendes Objekt zu greifen. Die Autoren präsentieren ein dafür geeignetes adaptives Regelungsschema, welches auf der Minimierung einer Gütefunktion basiert.

[Crétual et al. 98] extrahieren Bewegungsinformationen zur Schätzung der Lage eines komplexen Objekts im Bildbereich. Auf der Grundlage der berechneten Daten wird die auf einer Schwenk-Neige-Einheit montierte Kamera dem Objekt nachgeführt, um sicherzustellen, daß das beobachtete Objekt stets in der Mitte des Bildes zu sehen ist und sich insbesondere nicht aus dem Bild bewegt. Das Verfahren wird bei der Verfolgung einer Person eingesetzt.

[Ruf & Horaud 00] stellen einen Ansatz zur binokularen Regelung eines Roboters vor, der völlig im projektiven Raum formuliert ist. Metrische Modelle werden nicht verwendet. Stattdessen kommen projektive Modelle sowohl bei dem Stereokamerasystem als auch bei der Roboterkinematik zum Einsatz. Einige Probleme, die oft bei der Verwendung metrischer Modelle auftreten (z.B. lokale Minima, divergierende Regelung), werden durch diesen Ansatz von vorneherein vermieden.

[Yeasin & Chaudhuri 00] stellen eine Methode zur Programmierung eines Roboters vor, bei dem die zu erledigende Aufgabe von einem Menschen mehrmals vor einem Stereokamerasystem vorgeführt wird. Die dabei eingesetzte Kalman-Filter-basierte Verfolgung schätzt die Position farbiger Aufkleber auf den Fingern und den Handgelenken des Menschen über die Zeit hinweg. Aus den Verfolgungsdaten wird die Trajektorie für den Endeffektor des eingesetzten Roboters ermittelt. Um Singularitäten zu vermeiden und eine optimale Pfadgenerierung zu erzielen, werden die durchzuführenden Aktionen mehrmals über den Sichtsensor eingegeben (Verwendung eines Trajektorienbündels).

[Lee & Bailey-Van Kuren 00] beschreiben ein Modell zur automatischen Demontage gebrauchter Güter, das Benutzerinteraktionen sowie Rahmenbedingungen des zu demontierenden Objektes berücksichtigt. Das Modell stellt die wesentlichen Hilfsmittel zur Verfügung, um die jeweils nächste zu demontierende Komponente unter Berücksichtigung des Objektbauplans effizient zu bestimmen. Das Modell wird in einer Simulation unter Verwendung eines realen Objekts und visueller Eingabedaten getestet.

Zusammenfassend fällt im Bezug auf die hier anstehenden Fragestellungen vor allem auf, daß einige Verfolgungen in Echtzeit funktionieren. Um bei der in vorliegender Arbeit behandelten Regelung in Sachen Geschwindigkeit den Anschluß an den aktuellen Stand der Wissenschaft nicht zu verlieren, erscheint eine Beschleunigung der Verfolgung sinnvoll (vgl. Abschnitt 4.5.1).

Der Artikel [Crétual et al. 98] ist insbesondere in Bezug auf die Bemerkung ab Seite 2 von Interesse. [Lee & Bailey-Van Kuren 00] könnte sich im Hinblick auf die zukünftig geplante Demontage vieler Teile des Motorraumes als sehr hilfreich erweisen. Aufgrund der Zielsetzungen und Schwerpunkte bei der hier untersuchten Demontage von Altfahrzeugen im Vergleich zu oben zitierten Verfahren ergibt sich zunächst keine Veranlassung, eine gegenüber der Literatur veraltete Funktionsweise des hier untersuchten Regelungssystems anzunehmen.

2.4 Bewertungsaspekte

[Grimson 90] definiert die *Robustheit* einer Erkennung bzw. Lokalisation als das Verhalten, daß sich die Anpassungsgüte nur zögerlich vermindert bei einer Zunahme von Störungen in den Eingabedaten, bei einer Abnahme der relevanten Daten (= Bildmerkmale, die zum Objekt gehören) und bei einer Zunahme von irrelevanten Daten. Für eine hohe *Flexibilität* muß das Verfahren unter möglichst vielen unterschiedlichen Umständen der Robustheitsdefinition genügen. Insbesondere muß ein in Realweltszenen einsetzbares Verfahren auch korrekt arbeiten bei partiellen Verdeckungen, bei Anwesenheit von Stördaten aufgrund komplexer Umgebungen oder bei nur spärlich vorhandenen relevanten Daten. Außerdem gehört nach Meinung des Autors zur Flexibilität auch, daß ein Verfahren mit möglichst vielen Arten von Sensoren kompatibel ist (z.B. Radar, taktile oder bildgebende Sensoren).

[Chhabra & Phillips 98] unterscheiden 1:1-, 1:n- sowie n:1-Zuordnungen zwischen den Ergebnissen eines Erkennungsverfahrens und der Grundwahrheit. Sie berücksichtigen außerdem die Anzahl falscher Alarme sowie nicht erkannter Treffer. Zur Bewertung von Anpassungsergebnissen wird mittels dieser Werte eine Fehlerrate, eine Rate falscher Alarme sowie eine Detektionsrate definiert, welche wiederum zu einem Maß für die Gesamtkosten verrechnet werden.

[Wiedemann et al. 98] bewerten die Extraktion von Straßenmarkierungen mit einigen definierten Qualitätskriterien. Dazu wird jeweils ein Maß zwischen 0 und 1 für die Komplettheit, die Korrektheit sowie die Qualität eingeführt. Ferner wird eine Formel zur Berechnung der Redundanz (Zahlenwerte aus \mathbb{R}) angegeben sowie ein Maß für die Anzahl an Datenlücken pro Kilometer und die mittlere Lückenlänge.

[Heimes 00] beschreibt die Entwicklung eines Systems zur automatischen, sichtsystemgestützten Führung eines Fahrzeugs in Straßenverkehrsszenen. Der Autor stößt bei der Realisierung auf einige Bewertungsaspekte und bietet praktikable Lösungen an. Die *Zuverlässigkeit* des Regelungsmechanismus ist bei einem solchen System natürlich von besonderer Bedeutung. Aber auch der *Geschwindigkeitsaspekt* ist entscheidend für die zuverlässige Führung des Fahrzeuges. Die Verringerung der Zykluszeit des Bildauswertungsprozesses auf einen Wert unterhalb der Bildeinzugszeit ermöglicht die Auswertung eines jeden von den Kameras gelieferten Bildes (*Echtzeitauswertung*) und trägt entscheidend zur *Robustheit* der Regelung sowie zur *Sicherheit* des Systems bei. Durch eine Verbesserung der Latenzzeit wird unmittelbar die Reaktionsfähigkeit des Systems auf sich ändernde Rahmenbedingungen erhöht, was im Hinblick auf den *Sicherheitsaspekt* von besonderer Bedeutung ist. Zur Beschleunigung des entwickelten Verfolgungsprozesses wird eine Verteilung auf mehrere Rechner realisiert. Beispielsweise erfolgt die Darstellung der Ergebnisse auf einem anderen Rechner als die Berechnung der Verfolgungsergebnisse.

[Li et al. 98] bewerten bzw. vergleichen die vorgestellten Lokalisationsverfahren anhand dreier Gütekriterien: an der *Robustheit bzgl. einer Variation in den Ausgangsbedingungen*, der *Genauigkeit* der berechneten Ergebnisse sowie der *Effizienz* der Verfahren. Die Robustheit der Verfahren wird verglichen, indem unterschiedliche Teilmengen der relevanten Eingabedaten ausgewählt und den Algorithmen eingegeben werden. Die Anpassungsgenauigkeit ist definiert als das Tupel aus rotatorischem und translatorischem Fehler bei der Lageschätzung. Zur Bestimmung der Effizienz wird einfach die Lokalisationszeit gemessen.

Bei einigen Artikeln der Literatur wird eine mehr oder weniger intuitiv zugängliche Verdeutlichung der erzielten Robustheit angestrebt. Beispielsweise begründet [Rosin 99] die Robustheit des vorgestellten Lokalisationsverfahrens anhand der Resistenz gegenüber bis zu knapp 50% Ausreißern in den Eingabedaten. Zum Nachweis der Robustheit des vorgestellten Verfolgungsverfahrens wählen [Marchand et al. 99] einige als besonders schwierig erachtete Situationen bei der Verfolgung von Schraubenmuttern in Realwelt-Szenen aus.

Die in [Chhabra & Phillips 98] und [Wiedemann et al. 98] vorgestellten Gütekriterien umfassen zwar viele Details möglicher Fehlanpassungen, sind aber im Hinblick auf die hier anstehenden Untersuchungen nicht geeignet, da sie ohne Grundwahrheit kaum automatisch aus dem Bildmaterial zu bestimmen sind. Die übrigen Artikel geben wichtige Impulse bezüglich der hier anstehenden Definition von Genauigkeit und Robustheit sowie hinsichtlich Zuverlässigkeit, Flexibilität und Sicherheit eines Systems. Aufgrund der Funktionsweise der sichtsysteemgestützten Fahrzeugführung [Heimes 00] im Vergleich zu der automatischen Demontage ergeben sich in natürlicher Weise Parallelen hinsichtlich der Echtzeitfähigkeit und dem Bezug zur Zyklus- bzw. Latenzzeit sowie der Auswirkungen auf den Sicherheitsaspekt. Insbesondere dient auf beiden Gebieten die Bildeinzugszeit als kritische Schwelle für die Zykluszeit der Bildauswertung (vgl. Abschnitt 4.5.1).

Kapitel 3

Personenverfolgung

Wie in Kapitel 1 bereits angesprochen wurde, zeigt sich der in [Wachter 97] bzw. [Wachter & Nagel 99] dokumentierte Stand der Personenverfolgung prinzipiell einsetzbar. Nun geht es in erster Linie darum, das System so weit zu verbessern, daß man von einem in der Praxis tatsächlich anwendbaren Verfahren sprechen kann. Robustheitsfragen und Flexibilitätsfragen haben daher höchste Priorität. Demgegenüber treten zum gegenwärtigen Entwicklungsstand Aspekte hinsichtlich der Ausführungsgeschwindigkeit oder gar einer Echtzeitfähigkeit weit zurück. Auch die Schaffung einer Möglichkeit, die Verfolgung automatisch initialisieren zu können, muß an dieser Stelle aufgrund der Wichtigkeit der noch nicht hinreichend erreichten und damit als Erstes zu erzielenden Robustheit zurücktreten.

Der Schwerpunkt liegt also auf der Entwicklung eines flexiblen und robusten Verfahrens ungeachtet der dafür benötigten Rechenzeit (zumindest solange diese vernünftige Forschungen zuläßt). Folglich ergeben sich für die Weiterentwicklung des Verfahrens folgende zwei Schwerpunkte:

- Zum einen müssen die einzelnen Verfolgungsschwierigkeiten analysiert und reduziert werden, um die Zuverlässigkeit und Robustheit des Verfahrens zu steigern.
- Zum zweiten ist anstatt der in Abschnitt 1.1 beschriebenen bildfolgenabhängigen Parametrisierung eine bildfolgenunabhängige zu verwenden. Dies ist besonders wichtig, um das Verfahren allgemein auf unterschiedliches Bildmaterial anwenden zu können. Außerdem erleichtert dieser Punkt den Vergleich von Ergebnissen bei unterschiedlichen Bildfolgen und ermöglicht damit konkrete Schlußfolgerungen auf eventuelle Schwachstellen oder Fehlerursachen.

Beim zweiten Punkt ist a-priori damit zu rechnen, daß sich im Mittel bzgl. der Testbeispiele eine deutliche Verschlechterung der Verfolgung ergibt, schließlich ist anzunehmen, daß die von S. Wachter eingeführte Parametrisierung aus Kapitel 1 durch geeignete Optimierung der Verfolgungsergebnisse entstanden ist. Eine anschließende Bearbeitung des ersten Punktes würde dadurch erschwert. Zudem sind nach Abarbeitung des ersten

Punktes bessere Resultate bei Punkt 2 zu erwarten als bei alleiniger Durchführung des zweiten Punktes. Daher wird zuerst der erste Punkt in Angriff genommen (Abschnitt 3.1). Anschließend erfolgt in Abschnitt 3.2 die Realisierung der bildfolgenunabhängigen Parametrisierung.

Die Verwendung geeigneter Terminologie sowie die Bildung neuer Begriffe stellt in der Forschung das grundlegende Werkzeug für Argumentations- und Entwicklungsprozesse verschiedener Art bereit. Zuerst wird daher in Bezug auf obige Punkte die geeignete Terminologie definiert (Abschnitt 3.1). In ähnlicher Weise ist es sinnvoll, auch dem zugrundeliegenden Programmsystem einen aussagekräftigen Namen zu geben, schließlich sind die Untersuchungen damit durchzuführen. Da das der Personenverfolgung zugrundeliegende Programmsystem noch keinen Namen hat, wird es **Xtaro** getauft („tracking of articulated objects in image sequences“, eine Applikation unter **X**). Der Name betont die Fähigkeit des Systems, insbesondere *mehrgliedrige* Objekte verfolgen zu können.

3.1 Genauigkeit und Robustheit

Will man die Robustheit eines Verfolgungsverfahrens verbessern, so muß man natürlich zunächst festlegen, nach welchen Maßstäben man dies tun will. Folglich wird ein Gütekriterium für die Verfolgung in Form einer Definition der Robustheit benötigt. Zur Beurteilung der Anpassungsgenauigkeit ist zudem eine Definition einer solchen erforderlich. Damit können direkt die vorhandenen Verfolgungsschwierigkeiten identifiziert werden. Nach einer geeigneten Modifikation des Verfahrens kann anschließend das gleiche Gütekriterium dazu genutzt werden, um die Auswirkung der durchgeführten Modifikation auf die Verfolgungsergebnisse zu bewerten. Insbesondere ermöglicht das eingeführte Gütekriterium in natürlicher Weise einen Vergleich zwischen mehreren Verfahrensversionen oder Parametrisierungen und bildet so die Grundlage für eine schrittweise, systematische Verbesserung des Systems.

Nach einer Definition von Genauigkeit und Robustheit folgt im Anschluß die systematische Bearbeitung der einzelnen Problemfälle. Als Vorgehensweise hat sich dabei als sinnvoll erwiesen, die sich jeweils am stärksten zeigenden Verfolgungsprobleme bzw. Schwachstellen zuerst zu behandeln, da diese den besten Nährboden für eine erfolgreiche Analyse der zugrundeliegenden Ursache liefern. Wie sich herausstellen wird, sind die Problemursachen z.T. sehr unterschiedlicher Natur. Dies läßt die Möglichkeit einer einzelnen umfassenden Verfahrensmodifikation zur Verbesserung vieler Schwachstellen unwahrscheinlich werden und macht mehrere Modifikationen erforderlich.

3.1.1 Definition von Genauigkeit und Robustheit

Die *Genauigkeit* einer Modellanpassung ist bestimmt durch den *maximal auftretenden Anpassungsfehler*, welcher in dieser Arbeit mit Θ ($\Theta \in \mathbb{R}_{\geq 0}$) bezeichnet wird. Das Maximum des Anpassungsfehlers kann sowohl für einen Objektteil als auch für das

ganze Objekt in einer Aufnahme oder einer ganzen Bildfolge bestimmt werden. Man kann somit von der Anpassungsgenauigkeit eines oder mehrerer Objektteile oder des ganzen betrachteten Objekts sprechen und dies beziehen auf eine Aufnahme oder eine Bildfolge. Eine Anpassung bzgl. eines Bildes oder einer Bildfolge kann als *genau* oder *gut* bezeichnet werden, wenn der maximale Anpassungsfehler Θ unter einer zuvor vorgegebenen Schwelle Θ_S liegt. Ist die Grundwahrheit – d.h. die tatsächliche Lage des beobachteten Objekts in der realen Szene – bekannt, so ist dabei die Abweichung zwischen der aus dem Bild extrahierten Lageschätzung und der Grundwahrheit *in der Szene* geeignet zu bestimmen. Bei fehlender Grundwahrheit – und dies wird meist der Fall sein, da die Grundwahrheit üblicherweise schwer zu ermitteln ist – kann die maximale Abweichung zwischen dem ins Bild rückprojizierten Modell und den zugehörigen Bilddaten *im Bildbereich* ermittelt werden.

Es ergibt sich direkt folgende Unterscheidung.

- Die Anpassung zwischen projiziertem Modell und den Bilddaten ist im Bildbereich genau (bezüglich eines vorgegebenen Θ_S).
- Die tatsächliche Lage im Raum ist genau (also gut bezüglich der Grundwahrheit und einer vorgegebenen Genauigkeitsschwelle).

Diese beiden Bewertungen einer Anpassung sind voneinander abzugrenzen. Natürlich wird weitere Bewertung letztendlich angestrebt. Wenn das eingesetzte Bildauswertungsverfahren Sinn machen soll, muß die Erfüllung des ersten Punktes als Vorbedingung des zweiten angesehen werden. Ohne Kenntnis der Grundwahrheit ist die Erfüllung des ersten Punktes das Maximum dessen, was man tun kann. Tatsächlich wird sich später in der vorliegenden Arbeit zeigen, daß in der Erfüllung des ersten Punktes zunächst die eigentlichen Schwierigkeiten liegen. Es macht daher wenig Sinn, den zweiten Punkt anzugehen, bevor der erste zufriedenstellend gelöst ist. Im Falle einer Anpassung, die Punkt eins aber nicht Punkt zwei genügt, sind die verwendeten Kalibrierungen zwischen Szene und Welt, d.h. insbesondere die interne und externe Kamerakalibrierung, oder Modellierungsungenauigkeiten zu überprüfen und zu verbessern. Die Argumentation dieses Paragraphen überträgt sich direkt auch auf die weiter unten definierte Robustheit einer Anpassung.

Natürlich stellt sich bei fehlender Grundwahrheit die Frage, inwieweit Θ im Bildbereich automatisch vom eingesetzten Bildauswertungsrechner bestimmbar ist. Um einen entsprechenden Algorithmus zu kreieren, müßte man zu jedem Kantenelement im Bild zuverlässig berechnen können, ob es zum Objekt gehört oder nicht, welcher projizierten Bildkante und welcher Stelle darauf es zuzuordnen ist. Dies sind aber genau die bei der Verfolgung noch ungelösten Fragestellungen. Wenn man sie bei der Bewertung der Anpassungsgüte bereits hinreichend berechnen könnte, so hätte man daraus auch gleich einen genau (bzgl. eines kleinen Θ_S) funktionierenden Verfolgungsalgorithmus bauen können. Solange kein hinreichend funktionierendes Verfolgungsverfahren existiert, ist kein Automatismus zur Berechnung von Θ sinnvoll, d.h. Θ sollte manuell bestimmt

werden.¹

Die *Robustheit* einer Modellanpassung hängt ab von der Anpassungsgenauigkeit und den vom Bildmaterial gestellten Anforderungen, mit denen der Lokalisations- bzw. Verfolgungsmechanismus fertig wird. Ein Verfahren ist als umso robuster nachgewiesen, je kleiner der Anpassungsfehler Θ bzgl. der Bilddatenmenge und je umfangreicher, komplexer und verschiedenartiger diese Datenmenge ist. Mit Verschiedenartigkeit sind dabei unterschiedliche Objekte, Hintergründe und Konfigurationen in verschiedenen Bildern oder Bildfolgen gemeint. Unter Komplexität der Testdatenmenge fallen z.B. die Ausgeprägtheit der Objektkanten im Bild, die Beschaffenheit der Umgebung, d.h. der Störkanten und sonstiger störender Bildstrukturen, die Stärke auftretender Verdeckungen, der Bildkontrast, vorhandene Schlagschatten und Helligkeitsschwankungen, die Geschwindigkeit sich bewegender Objekte oder die Stärke des Rauschens im Bildmaterial. Wenn das betrachtete Lokalisations- bzw. Verfolgungsverfahren für eine Menge an Bildern oder Bildsequenzen Modellanpassungen liefert, die mindestens genau sind bezüglich eines Wertes Θ , so heißt das Verfahren für diese Testdaten *robust bezüglich Θ* oder kurz *Θ -robust*. Ähnlich wie bei der Anpassungsgenauigkeit kann auch hier eine Schwelle Θ_S vorgegeben werden, bis zu der man eine Anpassung als robust bezeichnen möchte. Wenn das Verfahren für das Testmaterial Θ -robust ist und $\Theta \leq \Theta_S$ gilt, so heißt das Verfahren (bzgl. der Testdatenmenge) auch kurz *robust*.

Ein Lokalisations- oder Verfolgungsverfahren, dessen Ergebnisse empfindlich von der Parametrisierung abhängen, wird bei hinreichend breit gewähltem Testmaterial in der Praxis kaum robust zum Laufen zu bringen sein. Daher spielt im Hinblick auf Robustheitsfragen auch die Abhängigkeit der Verfolgungsgüte von Veränderungen in der Parametrisierung eine Rolle.

Θ wird in dieser Arbeit stets in Bildpixeln angegeben. Bei der Festlegung von Θ ist die Auflösung des Bildmaterials zu berücksichtigen und ggf. auch die Größe des projizierten Modells im Bild. Durch einen Wert $\Theta < \infty$ wird ausgedrückt, wie genau oder ungenau bzw. wie robust die durchgeführte Objektverfolgung nach Beurteilung durch einen menschlichen Betrachter vollzogen wird. Mit $\Theta = \infty$ kann man dagegen zum Ausdruck bringen, daß ein Objektteil in der Modellprojektion an völlig falsche Bildstrukturen angepaßt wird und von einer Korrelation mit den zum Objektteil korrekterweise gehörenden Bildelementen nicht zu reden ist. Tritt während einer Bildfolge zu einem Objektteil der Übergang von $\Theta < \infty$ nach $\Theta = \infty$ auf, so spricht man auch: *‘das Objektteil wird verloren’*. Bei der Personenverfolgung gelte eine Verfolgung als robust bzw. genau, wenn $\Theta \leq \Theta_S := 3$ Pixel.

¹Dies belegen auch Experimente zur Verfolgung bei der sichtsystemgestützten automatischen Demontage, bei denen es nicht gelang, brauchbare Werte für Θ automatisch aus dem Bildmaterial abzuleiten.

3.1.2 Chaotischer Schwingungseffekt

Wie bereits zu Anfang dieses Unterkapitels erwähnt, ist es sinnvoll, die am deutlichsten ausgeprägten Verfolgungsprobleme zuerst anzugehen. Der sich am stärksten zeigende Anpassungsfehler findet sich zum gegebenen Stand der Entwicklung bei einem Versuch zur Bildfolge `alinde`. Ohne auf Anhieb erkennbaren Grund bewegt sich das gesamte projizierte Modell im Bild ab Halbbild 330 in nur 4 Halbbildern gegenüber den passenden Bilddaten zunächst ein Stück nach rechts und dann ein großes Stück (die halbe Breite der Person) nach links. Abbildung 3.1 verdeutlicht diesen Effekt.

Zur Analyse der zugrundeliegenden Ursache sind in Abbildung 3.2 die prädizierte Position der Person in x -Richtung (bzgl. des gewählten Weltkoordinatensystems) und die prädizierte Geschwindigkeit in x -Richtung der Verfolgung über der Zeit aufgetragen. Man kann erkennen, wie die Position nach einem ungefähr konstanten Verlauf schnell stark abfällt. Damit einher geht ein – ebenfalls nach etwa konstantem Verlauf – oszillierendes Verhalten der zugehörigen Geschwindigkeit, bei der sich die Amplitude in kurzer Zeit stark vergrößert. Aufgrund der sich dadurch ergebenden hohen Geschwindigkeiten muß die Personenverfolgung schließlich scheitern.

Die Ursache für das Problem kann zunächst wie folgt eingegrenzt werden. Abbildung 3.1 zeigt die prädizierten Zustände zu den Zeitpunkten 330 bis 333. Betrachtet man die aktualisierten Zustände, so kann man erkennen, daß diese zu den Zeitpunkten 330 bis 332 im Gegensatz zu den prädizierten zu guten Anpassungen bezüglich der Bilddaten gehören. Erst bei Halbbild 333 ergeben sich gravierende Abweichungen der aktualisierten Modellanpassung. Dies läßt sich wie folgt interpretieren: Die Ungenauigkeiten bei der Prädiktion können im Aktualisierungsschritt bis zum Zeitpunkt 332 ausgeglichen werden, da die Abweichungen zu den Bildkanten noch nicht zu groß sind. Bei Halbbild 333 liegt dagegen der Startwert für die Aktualisierung außerhalb des Fangbereichs des Minimierungsprozesses. Folglich kann als Grund für das Scheitern des Verfolgungsprozesses die Anpassung an die Bilddaten ausgeschlossen werden. Das Problem muß daher bei der Prädiktion liegen oder bei der Berechnung der Kovarianzmatrizen. Der Programmteil, der die Prädiktion in `xtaro` erledigt, wird geprüft. Da hier kein Fehler gefunden wird, ist die Berechnung der Kovarianzmatrizen zu prüfen.

In `xtaro` werden Modelle zweiter Ordnung bei der Freiheitsgradmodellierung eingesetzt, d.h. zu jedem zu schätzenden Wert wird auch seine zeitliche Änderung mitgeschätzt ([Wachter 97], S. 53/54). Der Systemzustand ξ besteht demnach aus einem Vektor \mathbf{u} , der die Positionen und Gelenkwinkel enthält, und einem Vektor \mathbf{v} , in welchem deren Änderungen enthalten sind. Die Kovarianzmatrix P läßt sich daher in vier Teile

$$P = \begin{pmatrix} P_{uu} & P_{uv} \\ P_{vu} & P_{vv} \end{pmatrix} \quad (3.1)$$

aufteilen mit $\xi^T P \xi = (\mathbf{u}^T \ \mathbf{v}^T) P \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \mathbf{u}^T P_{uu} \mathbf{u} + \mathbf{u}^T P_{uv} \mathbf{v} + \mathbf{v}^T P_{vu} \mathbf{u} + \mathbf{v}^T P_{vv} \mathbf{v}$.



Abbildung 3.1: Modellanpassungen zu einer Verfolgung zu `alinde` zu den Zeitpunkten 330 (links oben), 331 (rechts oben), 332 (links unten), und 333 (rechts unten). Dargestellt ist jeweils das tiefpaßgefilterte Bild mit dem angepaßten Modell.

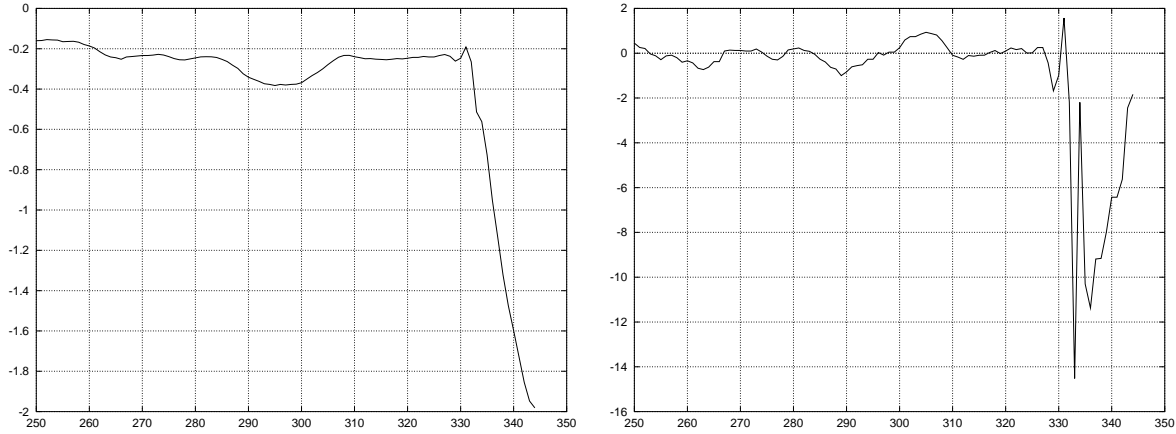


Abbildung 3.2: Prädierte Position der Person in x -Richtung (links) und Geschwindigkeit in x -Richtung (rechts) bei der Verfolgung zu `alinde` für den Zeitraum zwischen Halbbild 250 und dem Bildfolgende. Die Einheit der Ordinatenwerte ist m bzw. m/s .

Gemäß S. 52 in [Wachter 97] berechnet sich die prädierte Kovarianzmatrix \hat{P}_k^- zum Zeitpunkt k aus der aktualisierten Matrix $P := \hat{P}_{k-1}^+$ wie folgt ($\Delta t = t_k - t_{k-1}$):

$$\begin{aligned}
 \hat{P}_k^- &= \Phi_{k-1} P \Phi_{k-1}^T + \Gamma Q \Gamma^T \\
 &= \begin{pmatrix} I & \Delta t I \\ 0 & I \end{pmatrix} \begin{pmatrix} P_{uu} & P_{uv} \\ P_{vu} & P_{vv} \end{pmatrix} \begin{pmatrix} I & 0 \\ \Delta t I & I \end{pmatrix} + \Gamma Q \Gamma^T \\
 &= \begin{pmatrix} P_{uu} + \Delta t P_{uv} + \Delta t P_{vu} + \Delta t^2 P_{vv} & P_{uv} + \Delta t P_{vv} \\ P_{vu} + \Delta t P_{vv} & P_{vv} \end{pmatrix} + \Gamma Q \Gamma^T. \quad (3.2)
 \end{aligned}$$

Dabei beschreibt die Transitionsmatrix Φ_{k-1} den Übergang von Zeitpunkt t_{k-1} nach t_k und $\Gamma Q \Gamma^T$ die durch das Prozeßrauschen (normalverteilt, Spektraldichte Q) hinzukommende Unsicherheit.

In `Xtaro` werden intern nur die Matrizen P_{uu} , P_{uv} und P_{vv} gespeichert. Eine Kovarianzmatrix P hat im Programm stets die Gestalt

$$P = \begin{pmatrix} P_{uu} & P_{uv} \\ P_{uv}^T & P_{vv} \end{pmatrix}, \quad (3.3)$$

d.h. es wird angenommen, daß stets $P_{vu} = P_{uv}^T$ erfüllt ist. In `Xtaro` wird nicht die Matrix \hat{P}_k^- gemäß (3.2) berechnet, sondern stattdessen die Matrix

$$\bar{P}_k^- := \begin{pmatrix} P_{uu} + \Delta t P_{uv} + \Delta t P_{uv}^T + \Delta t^2 P_{vv} & P_{uv} + \Delta t P_{vv} \\ (P_{uv} + \Delta t P_{vv})^T & P_{vv} \end{pmatrix} + \Gamma Q \Gamma^T. \quad (3.4)$$

Es ist zu überprüfen, ob diese Berechnung richtig ist, d.h. ob $\bar{P}_k^- = \hat{P}_k^-$ gilt. Dies

Halbbild	$\ P_{uu} - P_{uu}^T\ $	$\ P_{vv} - P_{vv}^T\ $
30	0,000	0,000
150	0,000	0,000
220	0,000	0,000
250	0,000	0,000
270	0,000	0,000
310	0,000	0,000
318	0,000	0,042
320	0,000	0,258
322	0,001	1,334
324	0,003	7,846
326	0,019	48,147
328	0,117	291,520
330	0,674	1684,800
332	3,037	7593,090

Tabelle 3.1: Abweichungen der Matrizen P_{uu} und P_{vv} von ihren Transponierten bei der Verfolgung zu `alinde` zu einigen Halbbildzeitpunkten. Angegeben sind die Abweichungen der Teilmatrizen von $P = \bar{P}_k^-$.

ist dann der Fall, wenn $P_{vv} = P_{vv}^T$ gilt. Wenn auch $P_{uu} = P_{uu}^T$ gilt, so bedeutet das, daß $P = P^T$ ist (und das sollte bei Kovarianzmatrizen immer gelten). In diesem Fall gilt $\hat{P}_k^- = \bar{P}_k^- = \frac{1}{2}(\bar{P}_k^- + (\bar{P}_k^-)^T)$. In Tabelle 3.1 sind die tatsächlichen Abweichungen der Matrizen P_{uu} und P_{vv} von ihren Transponierten zu einigen Halbbildzeitpunkten aufgeführt. Aus der Tabelle kann man entnehmen, daß $P = P^T$ 318 Halbbilder lang hinreichend genau gilt. Ab Halbbildzeitpunkt 320 steigen die Abweichungen zwischen P und P^T exponentiell an. Diese Fehler setzen deutlich früher ein als bei der x -Position oder bei der Geschwindigkeit in x -Richtung, vgl. Abb. 3.2. Damit ist der Auslöser der Fehlanpassung zur Verfolgung mittels `alinde` gefunden. Den Gleichungen (3.2) bzw. (3.4) kann entnommen werden, wie sich ein Fehler in der Matrix P_{vv} auf die Matrizen P_{uv} und P_{uu} fortpflanzt. Tabelle 3.1 zeigt deutlich, wie sich dies im konkreten Versuch auswirkt. Eine Asymmetrie in der Kovarianzmatrix kann sich offenbar sehr schnell verstärken.

Nachdem der Auslöser für das Versagen des Verfolgungsprozesses gefunden ist, muß nun geklärt werden, wie eine Asymmetrie der Kovarianzmatrizen zustande kommen kann. Im Prädiktionsschritt (3.2) bzw. (3.4) kann höchstens die begrenzte rechnerinterne Berechnungsgenauigkeit für Gleitpunktzahlen die Symmetrie der Kovarianzmatrix stören, was zu vernachlässigen ist. Folglich muß es am Aktualisierungsschritt liegen, siehe Gleichungen (3.21) und (3.22) in [Wachter 97]. Bei Betrachtung der beiden Gleichungen

fällt sofort die Matrixinvertierung auf. Sie wird in `Xtaro` mittels des Gauß-Algorithmus mit Spaltenpivotsuche durchgeführt. Die Verteilung unvermeidbarer Berechnungsungenauigkeiten ist folglich i.A. asymmetrisch. Zudem ist bei dem Verfahren bekannt, daß insbesondere bei großen Matrizen signifikante Berechnungsungenauigkeiten auftreten können. Um die damit gefundene Fehlerquelle weitgehend auszuschließen, soll die Routine für die Matrixinvertierung ersetzt werden durch eine möglichst genaue Routine.

Nach eingehender Betrachtung einiger Literaturstellen ([Boehm & Prautzsch 93], [Engeln-Müllges & Reutter 90], [Press et al. 92], [Teubner 96]) und einer Implementation des Gesamtschrittverfahrens (vgl. [Teubner 96], Abschnitt 7.2.2), das sich aber als sehr langsam herausstellt, fällt die Wahl auf das Verfahren nach [Boehm & Prautzsch 93]. Eine $n \times n$ -Matrix A kann invertiert werden, indem die n linearen Gleichungssysteme $A\mathbf{x}_i = \mathbf{e}_i$, $i = 1, \dots, n$, gelöst werden. Hierbei bezeichnet \mathbf{e}_i den i -ten Einheitsvektor. Zur Lösung der linearen Gleichungssysteme wird die Methode der konjugierten Gradienten verwendet (vgl. [Boehm & Prautzsch 93], S. 34-37). Das Verfahren berechnet lineare Gleichungssysteme iterativ. Um möglichst genaue Ergebnisse zu erhalten, wird die Genauigkeit und damit die Iterationstiefe entsprechend hoch eingestellt. Die Personenverfolgung dauert bei Verwendung dieses Verfahrens insgesamt ca. 50% länger als beim Gauß-Algorithmus mit Spaltenpivotsuche. Zudem treten auch bei diesem Verfahren Ungenauigkeiten in der Symmetrie der sich ergebenden Kovarianzmatrizen auf. Die Möglichkeit eines sich aufschaukelnden Schwingungsverhaltens besteht also theoretisch weiter. Zur Lösung des Problems wird eine *Zwangssymmetrisierung* eingeführt, d.h. anstatt einer berechneten Inversen P wird $\frac{1}{2}(P + P^T)$ verwendet. Dies kann sowohl bei der ursprünglichen Invertiererroutine mit Gauß-Algorithmus als auch mit der Methode der konjugierten Gradienten eingesetzt werden. Versuche zeigen, daß das Problem in beiden Fällen gelöst ist. Die dabei erzielten Verfolgungsergebnisse sind in beiden Fällen gleich. Daher wird von nun an aus Geschwindigkeitsgründen der Gauß-Algorithmus mit Spaltenpivotsuche verwendet und mit der Zwangssymmetrisierung kombiniert eingesetzt.

Zusammenfassend kann man festhalten, daß der Verfolgungsprozeß bei der Personenverfolgung als rekursives System die Möglichkeit ungewünschter Schwingungseffekte in sich birgt, bei dem sich unvermeidbare Rechenungenauigkeiten zu makroskopischen Effekten aufschaukeln können. Am gegebenen Beispiel konnte das detailliert belegt werden. Daß dies kein Einzelfall ist, belegt eine (im Rahmen anderer Untersuchungen entsprechend parametrisierte) Verfolgung zu `hof_191096`, bei der ein ähnlicher Effekt zu beobachten ist. Ab Halbbild 188 entfernt sich ohne Zwangssymmetrisierung auch hier das gesamte projizierte Modell drastisch von den zugehörigen Bilddaten des untersuchten Fußgängers. Wie Tabelle 3.2 zeigt, tritt bei diesem Versuch der gleiche Effekt auf wie oben. Auch hier löst die Zwangssymmetrisierung das Problem.

Halbbild	≤160	161	168	178	181	184	186	187	188
$\ \hat{P}_k^- - (\hat{P}_k^-)^T\ $	0	10^{-6}	$1,3 \cdot 10^{-5}$	0,15	2,34	37,2	233	581	1436

Tabelle 3.2: Fehler in der Symmetrie der Kovarianzmatrizen \hat{P}_k^- zu einigen Halbbildzeitpunkten bei einer Verfolgung zu hof_191096.

3.1.3 Einsatz eines iterativ erweiterten Kalmanfilters unter Verwendung einer Levenberg-Marquardt–Optimierung

Als gravierendstes Problem bei der Personenverfolgung zeigt sich nach Beseitigung des chaotischen Schwingungseffektes die bei einigen Verfolgungen auftretende unzureichende Anpassung des aktualisierten Modells an die Bilddaten. In einigen Fällen macht die Anpassung der Arme oder Beine Schwierigkeiten, wenn der prädizierte Modellarm bzw. das prädizierte Modellbein zum Bildteil mehrere Pixel Abstand hat. Zudem finden sich zahlreiche Fälle ungenauer Anpassungen. Diese Probleme könnten darauf hindeuten, daß der Minimierungsprozeß im Aktualisierungsschritt nicht so funktioniert, wie er sollte, d.h. der Fangbereich zu klein ist oder lokale Minima der Gütefunktion erreicht werden. Um eine bestmögliche Anpassung sicherzustellen, wird der beste ermittelbare Algorithmus hierfür im folgenden eingesetzt.

Bei dem bisher in `xtaro` eingesetzten *erweiterten Kalman-Filter* (EKF) wird im Aktualisierungsschritt das *Gauß-Newton-Verfahren* zur Minimierung der verwendeten Gütefunktion verwendet (vgl. [Wachter 97], [Wachter & Nagel 99]). Abstiegsrichtung und Schrittweite werden durch Nullsetzen der Taylorreihe der Ableitung der Gütefunktion bestimmt

$$\nabla\Gamma(\mathbf{x} + \mathbf{s}) = \nabla\Gamma(\mathbf{x}) + \nabla^2\Gamma(\mathbf{x})\mathbf{s} + O(\|\nabla^3\Gamma(\mathbf{x})\| \cdot \|\mathbf{s}\|^2) \stackrel{!}{=} 0$$

nach einer Vernachlässigung der höheren Glieder:

$$\mathbf{s} = -(\nabla^2\Gamma(\mathbf{x}))^{-1}\nabla\Gamma(\mathbf{x}) . \quad (3.5)$$

Die Optimierung des Aktualisierungsschrittes läßt sich durch den Einsatz eines *iterativ erweiterten Kalman-Filters* (IEKF) verbessern. Bei diesem wird nicht nur in einer einmalig festgelegten Richtung ein Schritt in Richtung auf das Optimum zurückgelegt, sondern iterativ in mehreren Schritten bezüglich der dabei ermittelten günstigsten Richtungen \mathbf{s} . In jeder Iteration ist auch hier eine Optimierung der Bewertung mittels Gleichung (3.5) durchzuführen. An der Gleichung kann man unmittelbar die Hauptschwäche des Gauß-Newton-Verfahrens verdeutlichen: es ist die zweite Ableitung der Gütefunktion zu invertieren. Sollte $\text{rang}(\nabla^2\Gamma(\mathbf{x})) = 0$ auftreten, würde man dies ja noch direkt aufgrund einer Division durch Null an der vorzeitigen Terminierung der Verfolgung bemerken. Wesentlich kritischer ist aber der Fall einer schlecht konditionierten Matrix $\nabla^2\Gamma(\mathbf{x})$. In diesem Fall wird durch die Invertierung die ermittelte Abstiegsrichtung und damit die Minimierung fehlerhaft, was eine schlechte Anpassung

erwarten läßt, deren Ursache anhand der Anpassungsergebnisse nicht unmittelbar erschlossen werden kann. Um diese Möglichkeit auszuschließen, wird daher der iterativ erweiterte Kalman-Filter mit der *Levenberg-Marquardt-Optimierung* kombiniert. Vor der Invertierung von $\nabla^2\Gamma(\mathbf{x})$ wird bei der Berechnung von \mathbf{s} ein Term addiert, welcher die gute Konditionierung der Matrix sicherstellt. Die Optimierung realisiert eine Linearkombination von dem Gauß-Newton-Verfahren und dem Gradientenabstiegsverfahren, vgl. [Press et al. 92]. Bei guter Konditionierung von $\nabla^2\Gamma(\mathbf{x})$ wird das schnell konvergierende Gauß-Newton-Verfahren ausgeführt und bei schlechter Konditionierung das langsam konvergierende, aber dafür stabile Gradientenabstiegsverfahren.

Nach eingehender Betrachtung einiger Literaturstellen ([Gelb et al. 74], [Grewal/Andrews 93], [Daniilidis 92], [Koller 92], [Bar-Shalom & Fortmann 88]), wird der iterative erweiterte Kalmanfilter wie folgt implementiert:

Algorithmus Iterierter erweiterter Kalmanfilter unter Verwendung des

Levenberg-Marquardt-Verfahrens zur Minimierung der Güte Γ

Eingabe:

prädizierter Zustandsvektor $\hat{\mathbf{x}}^-$,

Kovarianzmatrix P^- (geht ein bei der Berechnung von $\overline{\nabla^2\Gamma}$ und $\nabla\Gamma$)

$\mu_0 := 10^{-6}$

$\nu := 10$

$\mu_{\max} = 10^9$

$k_{\max} := 20$

$\hat{\mathbf{x}}^+ := \hat{\mathbf{x}}^-$

$c := \Gamma(\hat{\mathbf{x}}^+) \quad /* \Gamma(\mathbf{x}) = \Gamma_{\hat{\mathbf{x}}^-}(\mathbf{x}) */$

$\mathbf{g} := \nabla\Gamma(\hat{\mathbf{x}}^+) \quad /* \text{Gradient von } \Gamma */$

$H := \overline{\nabla^2\Gamma}(\hat{\mathbf{x}}^+) \quad /* \text{approximierte Hessematrix von } \Gamma */$

$k := 0$

iteriere

$\mu := \mu_0$

iteriere

$\mathbf{s} := -(H + \mu I)^{-1}\mathbf{g} \quad /* \text{Abstiegsrichtung } (I = \text{Einheitsmatrix}) */$

$\mathbf{x}^* := \hat{\mathbf{x}}^+ + \mathbf{s}$

$\delta := \Gamma(\mathbf{x}^*) - c$

wenn $\delta \geq 0$

$\mu := \mu \cdot \nu$

sonst $/* \text{Verringerung von } \Gamma */$

$\hat{\mathbf{x}}^+ := \mathbf{x}^*$

```

c := Γ(x̂+)
g := ∇Γ(x̂+)
H := ∇2Γ(x̂+)
bis δ < 0 oder μ > μmax
k := k + 1
bis (δ ≥ 0 und μ > μmax) oder k > kmax
P+ := H-1 /* vgl. [Wachter 97], Seite 57 */

```

Ausgabe: Minimum $\hat{\mathbf{x}}^+$ von Γ (aktualisierter Zustandsvektor) und aktualisierte Kovarianzmatrix P^+ .

Es sei darauf hingewiesen, daß in der Schleife sichergestellt ist, daß die Gütefunktion auf jeden Fall verringert wird. Zudem wurde darauf Wert gelegt, die Terminierung des Verfahrens so weit wie möglich sicherzustellen und P^+ konform zu $\hat{\mathbf{x}}^+$, d.h. konform zum Kalman-Filter zu halten (bei der bisherigen Gauß-Newton-Minimierung wurde P^+ unabhängig davon bestimmt, wie weit bei der Minimierung tatsächlich in Abstiegsrichtung gelaufen wurde). Der Parameter k_{\max} kann über ein Menü in **xtaro** eingestellt werden, der oben im Algorithmus angegebene Wert hat sich allerdings bei der Verfolgung in zahlreichen Versuchen bewährt. Die genauen Werte μ_0 , ν und μ_{\max} sollten auf das Verfahren einen höchstens unerheblichen Einfluß haben, vgl. oben genannte Literaturstellen.

In [Wachter 97] wird auf Seite 78 erwähnt, daß der Rang der berechneten approximierten Hessematrix H beim Minimum der Gütefunktion Γ Null sein kann. Im Verlauf der Minimumsuche mittels des dargestellten Algorithmus könnte folglich eine zunehmende Verschlechterung der Kondition von H auftreten. Der dargestellte Algorithmus fängt zwar einige der damit verbundenen Probleme ab, jedoch ist zur Bestimmung der aktualisierten Kovarianzmatrix P^+ die Inverse von H zu berechnen. Dadurch würde sich eine schlechte Konditionierung in falschen Werten in P^+ und damit einer Störung der Verfolgung äußern. Daher ist es wichtig, zumindest exemplarisch das zeitliche Verhalten der Güte der Matrizen H zu studieren. Eine Untersuchung der auftretenden Konditionszahlen ist folglich sinnvoll. Aus diesem Grund sind in Abbildung 3.3 die kleinsten Eigenwerte und die Konditionszahlen für alle berechneten approximierten Hessematrizen zu einer Verfolgung mittels einer Bildfolge (**hof_191096**) dargestellt. Die ganzzahligen Abszissenwerte repräsentieren die Halbbildnummern. Die zu einer Halbbildnummer berechneten (mehreren) Werte sind jeweils ab dem entsprechenden ganzzahligen Abszissenwert aufgetragen. Man kann erkennen, daß sich die Kondition der berechneten Matrizen höchstens unwesentlich während des Minimierungsprozesses verschlechtert. Die Möglichkeit, daß der Rang von H beim Minimum von Γ Null werden kann, tritt offenbar bei der gegebenen Bildfolge nicht auf. Bei den anderen Bildfolgen ergibt sich kein Hinweis auf ein anderes Verhalten.

Von dem eingeführten IEKF unter Verwendung der Levenberg-Marquardt-Optimierung

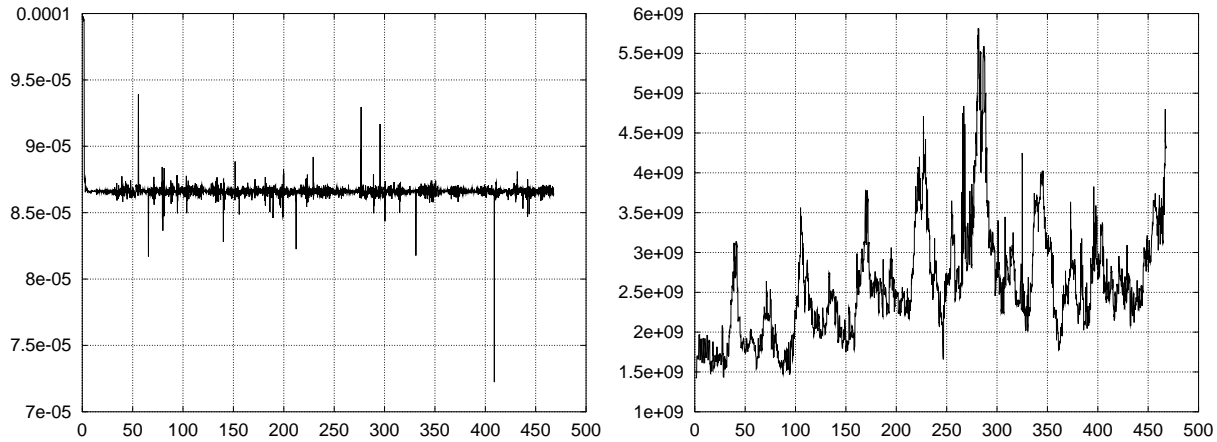


Abbildung 3.3: Kleinste Eigenwerte (links) und Konditionszahlen (rechts) der in den Aktualisierungsschritten zu einer Verfolgung mittels `hof_191096` auftretenden approximierten Hessematrizen. Die Werte zu einer Halbbildnummer sind jeweils ab dem entsprechenden Abszissenwert aufgetragen.

wird erwartet, daß sich die Verfolgungsergebnisse im Mittel verbessern im Vergleich zu vorher. Zur Überprüfung dieser Hypothese wird die Verfolgung anhand der zur Verfügung stehenden sechs Bildfolgen unter Verwendung des IEKF mit der neuen Minimierung durchgeführt und die Ergebnisse den entsprechenden Versuchen mit EKF und Gauß-Newton-Verfahren (Referenzversuche) gegenübergestellt. Tabelle 3.3 faßt die Unterschiede zusammen. Als Beispiel für solch einen Unterschied ist in Abbildung 3.4 der Unterschied in der Anpassung des rechten Beines bei `s1_191096` zum Halbbildzeitpunkt 209 dargestellt. Man kann an der Position des ins Bild projizierten Modells des rechten Unterschenkels erkennen, wie sich der linke Fuß im Bild beim Referenzversuch stärker auf die Fehlanpassung auswirkt als bei Versuch 6.

Es lassen sich beim Vergleich der Versuche 1 bis 6 mit den Referenzversuchen im wesentlichen zwei Klassen erkennen, in deren Zusammenhang signifikante Unterschiede bei der Modellanpassung auftreten. Zum einen treten häufig unterschiedliche Modellanpassungen eines Unterschenkels oder eines Beins auf im Zusammenhang mit Verdeckungen durch das jeweils andere Bein, vgl. Abb. 3.4. Zum anderen schaukeln sich ungenaue Anpassungen in einigen Fällen zu größeren Ungenauigkeiten auf, während in anderen Versuchen eine ähnliche, ebenfalls ungenaue Anpassung wieder korrigiert werden kann. Den auftretenden Anpassungen von Modellteilen an den Hintergrund gehen beispielsweise stets ungenaue Anpassungen an die zugehörigen Bildteile voraus. Das eigentliche Problem liegt hier also primär bei der ungenauen Anpassung und erst sekundär bei der anschließend auftretenden Anpassung an den Hintergrund.

In Tabelle 3.4 werden die Unterschiede der Versuche 1 bis 6 im Vergleich zu den Referenzversuchen kurz zusammengestellt. Im oberen Teil der Tabelle findet man die Unterschiede, die im Zusammenhang mit Beinverdeckungen stehen (oben genannte erste Klasse), im unteren Teil alle übrigen Unterschiede (obige zweite Klasse). Wie man

Vers.	Bildfolge	Bemerkungen
1	l_rot	Zwischen den Halbbildern 44 und 73 sowie zwischen 97 und 127 ergeben sich bessere Anpassungen des rechten Armes und im Zeitraum 333 bis 378 schlechtere Anpassungen des rechten Armes als im Referenzversuch. Ab Halbbild 365 und ab 397 erfolgen schlechtere Anpassungen des linken Unterschenkels jeweils aufgrund einer Verdeckung des linken Beins durch das rechte.
2	alinde	Ab 248 wird das rechte Bein schlechter angepaßt als im Referenzversuch. Aufgrund von Anpassungen des hinteren Rückenteils an den Hintergrund kommt es zwischen 268 und 298 zu einer schlechteren Anpassung des Rumpfes und des Beckens als im Referenzversuch. Zwischen 313 und 348 werden geringfügig bessere Anpassungen des Beckens, des Rückens und des rechten Oberschenkels beobachtet als im Referenzversuch.
3	frankm	Nach einer Verdeckung des linken Beins durch das rechte erfolgen jeweils in den Halbbildzeiträumen 68 bis 117 und 154 bis 175 bessere Anpassungen des linken Unterschenkels und in den Zeiträumen 128 bis 152 und 214 bis 315 schlechtere Anpassungen des linken Unterschenkels als im Referenzversuch.
4	g_braun	In den Zeiträumen 10 bis 58, 162 bis 237 sowie 330 bis 367 erfolgen bessere Anpassungen des linken Armes als im Referenzversuch. Zwischen 344 und 416 wird der Rumpf und der Kopf besser angepaßt und zwischen 413 und 494 wird die linke Schulter besser erfaßt.
5	hof_191096	Zwischen 12 und 28 erfolgen schlechtere Anpassungen des linken Unterschenkels als im Referenzversuch. Als sich der rechte Bildunterschenkel unter schwachem Kontrast zum Hintergrund aus der Verdeckung des linken Beins bewegt, werden zwischen 76 und 107 bessere Anpassungen des rechten Beins beobachtet als in Referenzversuch. Im Zeitraum 303 bis 326 erfolgen schlechtere Anpassungen des Rumpfes und des Beckens.
6	s1_191096	Im Vergleich zum Referenzversuch ergeben sich zwischen 208 und 215 bessere Anpassungen des rechten Beins (linkes Bildbein hat weniger Einfluß auf die Anpassung des rechten Modellbeins). Zwischen 319 und 338 erfolgen bessere Anpas-

Vers.	Bildfolge	Bemerkungen
		sungen des rechten Beins und ab 345 schlechtere als im Referenzversuch.

Tabelle 3.3: Versuche 1 bis 6 zur Minimierung von Γ mittels des iterierten erweiterten Kalmanfilters unter Verwendung der Levenberg-Marquardt-Minimierung. Unterschiede zu den Referenzversuchen: Zwangssymmetrisierungen, Minimierungsverfahren.

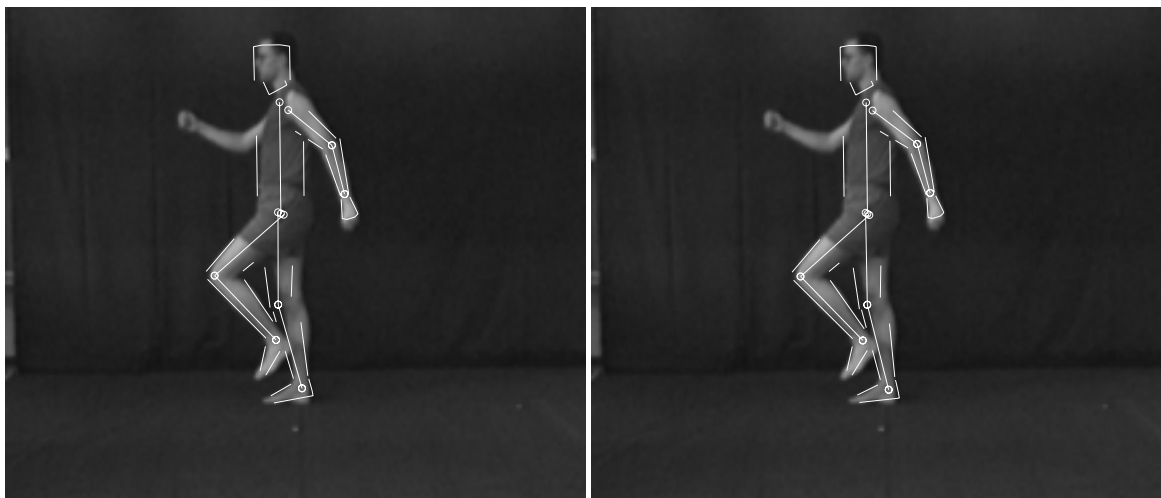


Abbildung 3.4: Modellanpassungen des Referenzversuchs (links) und bei Versuch 6 (rechts) zum Halbbildzeitpunkt 209. Dargestellt ist jeweils das tiefpaßgefilterte Bild mit dem präzidierten Modell.

erkennen kann, ergeben sich bei den Versuchen 1 bis 6 für *beide* Klassen mehr Verbesserungen gegenüber den Referenzversuchen als Verschlechterungen. Insgesamt zählt man 13 Verbesserungen und nur 10 Verschlechterungen im Vergleich zur Referenz. Folglich ist der iterativ erweiterte Kalmanfilter mit Levenberg-Marquardt-Optimierung als für die Verfolgung geeigneter nachgewiesen als das ursprüngliche Verfahren.

Man kann abschließend folgendes zusammenfassen:

- Keines der beiden verwendeten Minimierungsverfahren liefert ausschließlich bessere Ergebnisse als das andere.
- Bei beiden Verfahren treten ungenaue bzw. schlechte Anpassungen auf.
- Während eine ungenaue Anpassung in einigen Fällen zu immer schlechteren Resultaten führt, kann eine geringfügig abweichende Ausgangssituation in anderen Versuchen zu guten Ergebnissen führen.

Situation, Modellteil	
frankm, 68-117, linker Unterschenkel	besser
frankm, 154-175, linker Unterschenkel	besser
hof_191096, 76-107, rechtes Bein	besser
s1_191096, 208-215, rechtes Bein	besser
l_rot, ab Hb. 365, linker Unterschenkel	schlechter
l_rot, ab Hb. 397, linker Unterschenkel	schlechter
frankm, 128-152, linker Unterschenkel	schlechter
frankm, 214-315, linker Unterschenkel	schlechter
l_rot, 44-73, rechter Arm	besser
l_rot, 97-127, rechter Arm	besser
alinde, 313-348, Rumpf und Becken	besser
g_braun, 10-58, linker Arm	besser
g_braun, 162-237, linker Arm	besser
g_braun, 330-367, linker Arm	besser
g_braun, 344-416, Rumpf und Kopf	besser
g_braun, 413-494, linke Schulter	besser
s1_191096, 319-338, rechtes Bein	besser
l_rot, 333-378, rechter Arm	schlechter
alinde, ab Hb. 248, rechtes Bein	schlechter
alinde, 268-298, Rumpf	schlechter
hof_191096, 12-28, linker Unterschenkel	schlechter
hof_191096, 303-326, Rumpf u. Becken	schlechter
s1_191096, ab Hb. 345, rechtes Bein	schlechter

Tabelle 3.4: Zusammenstellung aller wesentlichen Unterschiede der Ergebnisse der Versuche 1 bis 6 im Vergleich zu den Referenzversuchen. Im oberen Teil der Tabelle sind die Unterschiede im Zusammenhang mit Beinverdeckungen aufgeführt, im unteren Teil alle übrigen Unterschiede.

Folglich hängen die verbleibenden Verfolgungsprobleme nicht vom verwendeten Optimierungsverfahren ab. Es muß daher weitere Ursachen für Robustheitsmängel geben.

3.1.4 Freiheitsgradwertebeschränkungen

Der nach dem gegebenen Stand der Dinge stärkste Anpassungsfehler findet sich bei der Verfolgung zu frankm bei Halbbild 282. Während bei Halbbild 281 die Anpassung des linken Beins genau durchgeführt wird, tritt ein Halbbild später plötzlich ein starkes



Abbildung 3.5: Modellanpassungen bei **frankm** zu den Halbbildzeitpunkten 281 (links) und 282 (rechts). Das linke Modellbein bewegt sich plötzlich unerwartet rückwärts trotz fortgesetzter Vorwärtsbewegung des linken Bildbeins. Dargestellt sind die tiefpaßgefilterten Halbbilder mit dem angepaßten Modell.

Zurückzucken des Modellbeins trotz fortgesetzter Vorwärtsbewegung des Bildbeins auf, siehe Abbildung 3.5.

Zwar ähnelt der auftretende Effekt dem in Abschnitt 3.1.2 beschriebenen, die Ursache ist jedoch eine andere, denn der Symmetriefehler $\|\hat{P}_k^- - (\hat{P}_k^-)^T\|$ von \hat{P}_k^- ist während der ganzen Verfolgung 0.

Bei genauerer Analyse der von **Xtaro** durchgeführten Modellanpassung findet sich die folgende Ursache des Problems. Die in Abschnitt 1.1 angesprochene mögliche Beschränkung von Freiheitsgraden wird bei dieser Verfolgung für den Beugewinkel des linken Knies verwendet. Die Extremwerte für diesen Winkel sind eingestellt auf $-0,15 \text{ rad}$ und $2,5 \text{ rad}$.² Bei der Prädiktion von Halbbild 280 auf 281 ergibt es sich nun, daß der Beugewinkel des linken Knies zu $-0,1502 \text{ rad}$ berechnet wird. Die Winkelgeschwindigkeit wird bestimmt zu $-7,3619 \text{ rad/s}$, siehe auch den oberen Teil von Tabelle 3.5. Der Wert $-0,1502 \text{ rad}$ für den Beugewinkel wird als zu klein erkannt und auf $-0,15 \text{ rad}$ gesetzt (unerhebliche Änderung des Wertes). Gleichzeitig wird die Beugegeschwindigkeit von $-7,3619 \text{ rad/s}$ auf 0 rad/s gesetzt, wodurch sich für die aktualisierte Geschwindigkeit zum Zeitpunkt 281 schließlich der Wert $2,1303 \text{ rad/s}$ ergibt, was etwa gleich der Korrektur der Geschwindigkeit ist, die sich ohne die Winkelbeschränkung ergeben hätte. Ohne die Freiheitsgradrücksetzung bei Halbbild 280 hätte sich bei Halb-

²Mit *rad* sei der Winkel im Bogenmaß bezeichnet.

Halbbild Berechnungsschritt	280		281		282	
	Präd.	Akt.	Präd.	Akt.	Präd.	Akt.
Beugen linkes Knie (rad)	-0,0104	-0,0030	-0,1500	-0,1068	-0,0642	-0,0479
Beugegeschw. (rad/s)	-7,8208	-7,3619	0,0000	2,1303	2,1303	2,8435

Halbbild Berechnungsschritt	280		281		282	
	Präd.	Akt.	Präd.	Akt.	Präd.	Akt.
Beugen linkes Knie (rad)	-0,0104	-0,0030	-0,1502	-0,1075	-0,2126	-0,1588
Beugegeschw. (rad/s)	-7,8208	-7,3619	-7,3619	-5,2549	-5,2549	-2,6575

Tabelle 3.5: Beugewinkel des linken Knies und dessen Winkelgeschwindigkeit bei **frankm**, wenn die Beschränkung des Winkels ab Halbbild 280 eingeschaltet (oben) bzw. abgeschaltet ist (unten). Dargestellt sind die Werte für die Prädiktions- (abgekürzt mit Präd.) und Aktualisierungsschritte (Akt.) zu den Halbbildzeitpunkten 280 bis 282.

bild 281 der aktualisierte Wert $-5,2549 rad/s$ ergeben (gleich $-7,3619 rad/s$ plus der Korrektur), vgl. unteren Teil der Tabelle 3.5. Dieser Wert kann als Näherung für die tatsächliche Änderung des Beugewinkels des linken Bildknies angesehen werden, da die Verfolgung bei Halbbild 282 robust bleibt, wenn die Beschränkung bei Halbbild 280 nicht stattfindet. Die Differenz zwischen geschätzter und tatsächlicher Winkelgeschwindigkeit ist damit zum einen relativ groß und zum anderen haben die Werte unterschiedliche Vorzeichen. Während sich der Bildunterschenkel vorwärtsbewegt, bewirkt folglich die Prädiktion auf Halbbild 282 eine auffallende, signifikante Rückwärtsbewegung des Modellunterschenkels.

Dies zeigt, daß die Beschränkung der Freiheitsgradwerte grundsätzlich auch zu Instabilitäten bei der Objektverfolgung führen kann. Im vorliegenden Fall kann das Problem beseitigt werden, indem gezielt die Beschränkung für den Beugewinkel des linken Kniegelenkes bei Halbbild 280 ausgeschaltet wird. Wenn die Beschränkung dieses Winkels für die gesamte Bildfolge abgeschaltet wird, so wird erwartungsgemäß eintreten, daß die gewonnene Stabilisierung der Verfolgung an anderen Stellen wegfällt, schließlich wurde die Beschränkung dieses Wertes gezielt zur Stabilisierung eingeführt.

Zur Behebung des Problems wird untersucht, ob durch Verändern des Extremwertes $-0,15 rad$ für den Beugewinkel des linken Knies eine Verbesserung erzielt werden kann. Zwischen Halbbild 38 und 41 erfolgen Beschränkungen des Knie-Freiheitsgrades. Diese erfolgen, weil der prädizierte Beugewinkel die Werte $-0,1675 rad$, $-0,1650 rad$, $-0,2196 rad$ bzw. $-0,1715 rad$ annimmt. Zu den zugehörigen Zeitpunkten bewegt sich der Bildunterschenkel mit kleiner Geschwindigkeit, weshalb sich hier das obige Problem nicht ergibt. Wird der Extremwert größer als $-0,1650 rad$ gewählt, so sollten sich ab Halbbild 38 höchstens kleine Veränderungen ergeben. Wird der Wert auch kleiner als $-0,1502 rad$ gewählt, so bleibt die Beschränkung bei Halbbild 280 aus. Dann sollte

Vers.	Extremwert	Bemerkungen
7	$-0,1576 \text{ rad}$	Das im Text genannte Problem bei Halbbild 282 tritt nicht mehr auf, da der Beugewinkel des linken Knies bei 281 völlig anders eingeschätzt wird als im Referenzversuch (Grund: Abweichungen ab Halbbild 38).
8	$-0,1503 \text{ rad}$	Im Gegensatz zum Referenzversuch wird das linke Bein bei Halbbild 92 verloren, obwohl der verwendete Extremwert nur $0,0003 \text{ rad}$ von dem aus dem Referenzversuch abweicht.

Tabelle 3.6: Versuche 7 und 8 zu `frankm` unter Verwendung der angegebenen Extremwerte anstatt $-0,15 \text{ rad}$ für den Beugewinkel des linken Knies.

aber eine Beschränkung bei Halbbild 282 stattfinden, vgl. Tabelle 3.5 (unterer Teil). Der Wert $-0,1576 \text{ rad}$ liegt zwischen den genannten Werten. Mit ihm soll untersucht werden, ob sich ab Halbbild 282 Instabilitäten ergeben (Versuch 7). Tabelle 3.6 faßt im oberen Teil die Unterschiede der beobachteten Ergebnisse im Vergleich zum Referenzversuch zusammen.

Aufgrund der Tatsache, daß in Versuch 7 ab Halbbild 38 bei den durchgeführten Beschränkungen der Beugewinkel auf einen anderen Wert gesetzt wird als im Referenzversuch, ergeben sich ab Halbbild 38 Modellanpassungen, die vom Referenzversuch zunächst geringfügig abweichen. Im weiteren Verlauf ergeben sich teilweise etwas größere Abweichungen, die Personenverfolgung bleibt jedoch bei beiden Versuchen ähnlich genau. Es sei angemerkt, daß die Beschränkungen des betrachteten Freiheitsgrades in Versuch 7 zu den Halbbildzeitpunkten 38, 39, 107 und 163 und nicht wie im Referenzversuch zu den Zeitpunkten 38, 39, 40 und 41 stattfinden. Bei Halbbild 281 führen die Abweichungen schließlich dazu, daß sich in Versuch 7 das Problem des Referenzversuchs gar nicht mehr stellt, denn der Beugewinkel des linken Knies wird völlig anders eingeschätzt, wie im oberen Teil von Abbildung 3.6 verdeutlicht wird. Dafür stellt sich ein anderes Verfolgungsproblem bezüglich des betrachteten Beins. Soll das verhindert werden, kann versucht werden, den Extremwert für den Beugewinkel so zu wählen, daß ab Halbbild 38 möglichst geringe Abweichungen zum Referenzversuch produziert werden. Gleichzeitig soll aber bei Halbbild 280 keine Beschränkung stattfinden. Es ist also der Extremwert $-0,1503 \text{ rad}$ zu wählen. In Versuch 8 wird die Modellanpassung für diesen Wert durchgeführt. In Tabelle 3.6 ist das Verfolgungsergebnis aufgeführt.

Obwohl die Verfolgung für die Werte $-0,1500 \text{ rad}$ und $-0,1576 \text{ rad}$ vor Halbbild 281 robust bleibt (Referenzversuch und Versuch 7), ergibt sich in Versuch 8 für den Wert $-0,1503 \text{ rad}$ dazwischen verblüffenderweise eine besonders deutliche Fehlanpassung bei der Verfolgung des linken Beines: Ab Halbbild 92 bleibt das linke Modellbein am rechten Bildbein hängen, während das rechte Bein das linke verdeckt. Das linke Bein wird

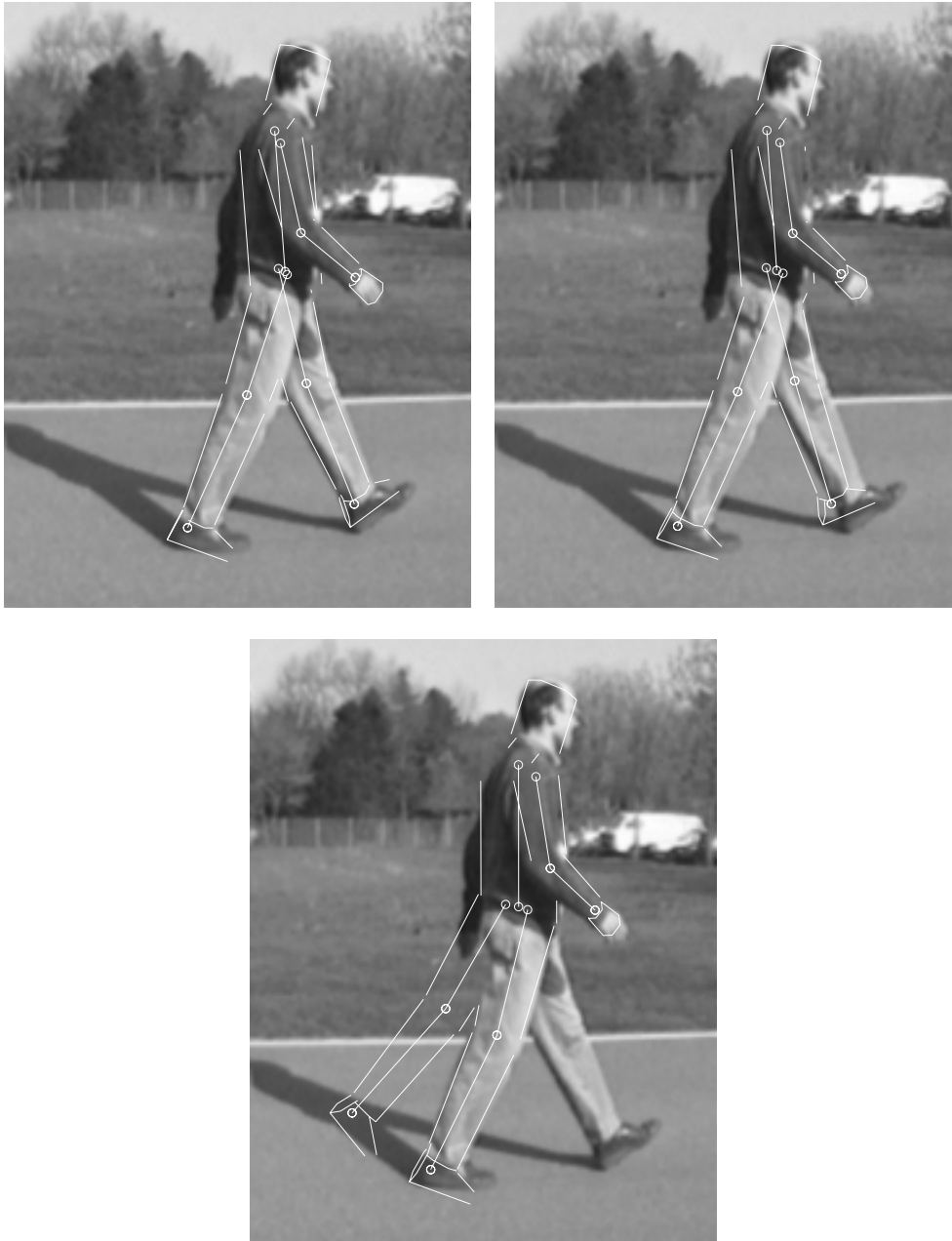


Abbildung 3.6: Modellanpassungen des Referenzversuchs (oben links) sowie bei Versuch 7 (oben rechts) und 8 (unten) bei **frankm** zum Halbbildzeitpunkt 281. In Versuch 7 ergibt sich das Problem des Referenzversuchs nicht mehr. In Versuch 8 wird das linke Bein verloren, obwohl der verwendete Extremwert nur $0,0003 \text{ rad}$ von dem aus dem Referenzversuch abweicht. Dargestellt sind jeweils die tiefpaßgefilterten Halbbilder mit dem angepaßten (prädizierten) Modell.

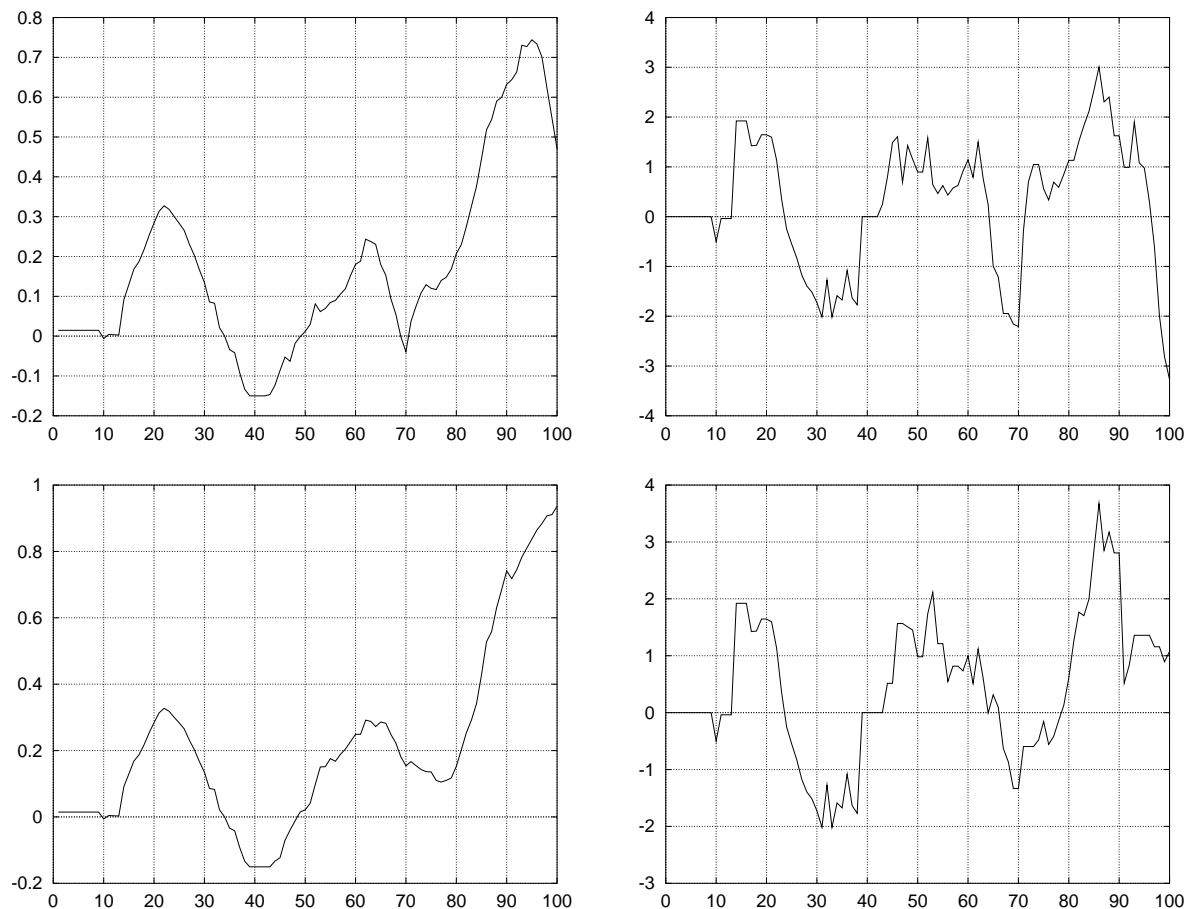


Abbildung 3.7: Prädizierter Winkel für die Beugung des linken Knies (links) und die zugehörige (prädierte) Winkelgeschwindigkeit (rechts) im Referenzversuch (oben) und in Versuch 8 (unten) zu *frankm*. Dargestellt ist der Zeitraum zwischen den Halbbildern 0 und 100. Die Einheit der Ordinatenwerte ist *rad* bzw. *rad/s*.

im weiteren Verlauf verloren, so daß es bei Halbbild 281 der Person ‘hinterher geschleift’ wird, vgl. Abb. 3.6. In Versuch 8 finden die Beschränkungen des betrachteten Freiheitsgrades wie auch im Referenzversuch zu den Zeitpunkten 38, 39, 40, 41 statt. Erst nachdem das linke Bein verloren wurde, erfolgen weitere Beschränkungen. Zur Illustration der zeitlichen Entwicklung des Beugewinkels des linken Knies sowie der zugehörigen Winkelgeschwindigkeit sind in Abbildung 3.7 zum Referenzversuch bzw. zu Versuch 8 die entsprechenden Werte zwischen den Halbbildern 0 und 100 über der Zeit aufgetragen.

Zusammenfassend zu den bisherigen Untersuchungen zu den Freiheitsgradwertebeschränkungen ist folgendes festzuhalten:

- Durch die Verwendung der Freiheitsgradwertebeschränkungen nach [Wachter 97]

können signifikante Störungen der Verfolgung auftreten.

- Auf die Freiheitsgradwertebeschränkungen kann allerdings nach dem gegenwärtigen Stand nicht verzichtet werden, schließlich wurden die Beschränkungen von S. Wachter gezielt zur Stabilisierung der Verfolgung eingeführt.
- Die Versuche zeigen, daß die Verfolgung ausgesprochen empfindlich vom gewählten Extremwert der Freiheitsgradrücksetzung abhängt. Es ist keinesfalls sichergestellt, daß sich ein ähnliches Verfolgungsverhalten ergibt, wenn man eine Freiheitsgradrücksetzung geringfügig ändert. Ein Verfahren, das so empfindlich von den Steuerparametern abhängt, wird bei hinreichend breit gewähltem Testmaterial in der Praxis kaum robust sein.

Folglich ergeben sich unmittelbar zwei Möglichkeiten zum weiteren Vorgehen:

- Es ist zu untersuchen, ob die Freiheitsgradwertebeschränkungen durch eine Modifikation so verbessert werden können, daß die genannten Schwierigkeiten ausbleiben oder zumindest gemildert werden.
- Wenn damit nicht die gewünschten Ergebnisse erzielt werden können, kann nur die Verbesserung des grundlegenden Verfolgungsalgorithmus (mit oder ohne Freiheitsgradwertebeschränkungen) zum Ziel führen.

Zum ersten Punkt kann festgestellt werden, daß das eingangs dokumentierte Verfolgungsproblem dadurch entstanden ist, daß außer dem Gelenkwinkel selbst auch die geschätzte Winkelgeschwindigkeit zurückgesetzt wurde. Läßt man die Rücksetzung der Winkelgeschwindigkeit weg, so wird erwartet, daß sich bei einer Beschränkung während der Vorwärtsschwingphase eines richtig verfolgten Beins (Referenzversuch zu frankm, Halbbild 280) keine Probleme ergeben. Befindet sich das richtig verfolgte Bein vor einem Umkehrpunkt der Bewegung (Referenzversuch zu frankm ab Halbbild 38), so dürfte sich auch hier kein Problem ergeben, denn der Aktualisierungsschritt sorgt für die betragsliche Reduzierung der Geschwindigkeit, d.h. eine Rücksetzung der Geschwindigkeit auf 0 rad/s sollte überflüssig sein. Findet eine Rücksetzung statt, wenn ein Objektteil bereits verloren wurde, so dürfte diese im Allgemeinen höchstens zufällig ein Wiedereinfangen des Teils zur Folge haben. Voraussichtlich ergeben sich also keine Probleme, wenn die Geschwindigkeitsrücksetzung unterbleibt.

Um die tatsächlichen Auswirkungen zu überprüfen, werden in den Versuchen 9 bis 14 die Versuche 1 bis 6 wiederholt ohne Rücksetzung der Geschwindigkeitsrücksetzung. Tabelle 3.7 faßt die auftretenden Unterschiede zusammen. Man zählt insgesamt 22 Fälle, bei denen in den Versuchen 9 bis 14 schlechtere Anpassungen auftreten und nur 7 Fälle besserer Anpassungen als in den Versuchen 1 bis 6. Daraus muß gefolgert werden, daß im Falle eingesetzter Freiheitsgradwertebeschränkungen die Rücksetzung der Geschwindigkeit zur Robustheit der Verfolgung beiträgt.

Vers.	Bildfolge	Bemerkungen
9	l_rot	Zwischen 248 und 265 sowie zwischen 427 und 447 ergeben sich im Vergleich zu Versuch 1 schlechtere Anpassungen des linken Unterschenkels nach einer Verdeckung durch den rechten Unterschenkel und zwischen 384 und 561 schlechtere Anpassungen der rechten Schulter und des rechten Oberarmes.
10	alinde	Im Vergleich zu Versuch 2 erfolgen zwischen 107 und 128 schlechtere Anpassungen des rechten Unterschenkels und zwischen 170 und 190 tritt eine schlechtere Einschätzung der Lage des Beckens auf. Zwischen 218 und 245 erfolgen wesentlich schlechtere Anpassungen des Beckens und des Rückens als in Versuch 2. Ab 248 ergeben sich bessere Anpassungen des rechten Beins und ab 268 schlechtere. In den Zeiträumen 146 bis 184, 243 bis 271 und 294 bis 324 erfolgen schlechtere Anpassungen des linken Oberarmes.
11	frankm	In den Zeiträumen 94 bis 127, 93 bis 117 sowie 155 bis 184 treten schlechtere Anpassungen und zwischen 128 und 152 bessere Anpassungen des linken Unterschenkels auf als in Versuch 3. Aufgrund von Freiheitsgradwertebeschränkungen des rechten Kniewinkels erfolgen zwischen 253 und 255 bessere Anpassungen des rechten Unterschenkels und in den Zeiträumen 256 bis 291 und 313 bis 408 schlechtere Anpassungen des rechten Unterschenkels. Zwischen 323 und 355 wird der rechte Arm schlechter angepaßt als in Versuch 12.
12	g_braun	Verfolgung ähnlich wie in Versuch 4.
13	hof_191096	Im Gegensatz zu Versuch 68 wird der rechte Unterschenkel ab Halbbild 76 nicht verloren. Der in Versuch 68 ab 162 auftretende Verlust der Bildperson tritt nicht auf. Nach einer Verdeckung paßt zwischen 288 und 297 das linke Modellbein schlechter zum Bildbein als in Versuch 5, zwischen 302 und 319 erfolgt im Vergleich zu Versuch 5 eine bessere Anpassung des Rumpfes und Beckens, zwischen 374 und 391 wird das rechte Bein schlechter angepaßt (nach Verdeckung) und zwischen 446 und 457 ergibt sich eine schlechtere Anpassung des rechten Unterschenkels offenbar aufgrund einer Beschränkung des Kniegelenk-Freiheitsgrades.
14	s1_191096	Im Vergleich zu Versuch 6 werden zwischen 194 und 199 sowie zwischen 208 und 215 schlechtere und zwischen 203

Vers.	Bildfolge	Bemerkungen
		und 206 bessere Anpassungen des rechten Beins beobachtet. Ab 340 bleibt der linke Arm im Gegensatz zu Versuch 6 stehen, als der Bildarm beginnt, sich zu bewegen.

Tabelle 3.7: Wiederholung der Versuche 1 bis 6 ohne Geschwindigkeitsrücksetzung bei den Freiheitsgradwertebeschränkungen.

Zusammenfassend ist Folgendes zu bemerken.

- Es wurde nachgewiesen, daß die Güte der Verfolgung sehr empfindlich (zu empfindlich im Bezug auf Robustheitsfragen) von den Werten bei den Freiheitsgradwertebeschränkungen abhängt. Insbesondere treten zu starke Änderungen der Verfolgungsgüte bei kleinen Änderungen der Werte auf.
- Durch die Freiheitsgradwertebeschränkungen treten signifikante Verfolgungsdefizite auf. Gleichzeitig stabilisieren sie die Verfolgung an anderer Stelle deutlich.
- Die Modifikation der Freiheitsgradwertebeschränkungen führte nicht zum Ziel.

Daraus folgt, daß das Hauptproblem bei der Verfolgung nicht in Bezug auf die Freiheitsgradwertebeschränkungen zu suchen ist, sondern an anderer Stelle. Die hauptsächliche Schwachstelle ist noch zu identifizieren. Wünschenswert wäre ein Verfolgungsverfahren, welches so robust funktioniert, daß auf Freiheitsgradwertebeschränkungen möglichst ganz verzichtet werden kann. Falls in Zukunft dennoch die Freiheitsgradwertebeschränkungen weiter benötigt werden sollten, ist als Ausblick und Anregung für die Zukunft zu beachten, daß die Freiheitsgradwertebeschränkungen *nicht Kalman-Filterkonform* integriert sind. Im Verlauf der Verfolgung werden direkt die Schätzgrößen massiv manipuliert. Ein solcher Eingriff muß den sensibel auf Störungen reagierenden (vgl. Abschnitt 3.1.2!) rekursiven Kalman-Filter-Prozeß signifikant negativ beeinflussen. Der Artikel [Geeter et al. 97] gibt hierzu möglicherweise eine wichtige Hilfestellung. Dieser Themenkomplex soll an dieser Stelle nicht bearbeitet werden, da nach obiger Schlußfolgerung die Hauptschwäche bei der Personenverfolgung an anderer Stelle zu suchen ist.

3.1.5 Zusammenfassung und Ausblick

Die vorausgehenden Abschnitte belegen einen deutlichen Fortschritt hinsichtlich der Robustheit bei der Personenverfolgung und zeigen gleichzeitig, wie vielschichtig und mühselig der Prozeß der Robustheitssteigerung ist. Schritt für Schritt müssen die Schwachstellen lokalisiert, analysiert und verbessert werden. Aufgrund der bisherigen Untersuchungen lassen sich einige verbleibende grundsätzliche Schwächen feststellen.

Diese sollen nun kurz umrissen werden, denn – auch wenn unmittelbar hier keine Lösungen angeboten werden können – so bildet natürlich die Erkenntnis der Schwachstellen oftmals den ersten Schritt zur Lösung. Folgendes mag daher als Ausblick für weitere Forschungen dienen.

Das Optimierungsverfahren im Aktualisierungsschritt konnte deutlich verbessert werden. Diesbezüglich zu beachten bleibt, daß es sich nach wie vor um ein lokal operierendes Verfahren handelt. Es besteht folglich prinzipiell die Möglichkeit, ein lokales Minimum anstatt des gewünschten globalen zu erhalten (vorausgesetzt, die Gütefunktion entspricht den Anforderungen, d.h. die Minima der Gütefunktion repräsentieren tatsächlich korrekte Lageschätzungen).

Die Abschnitte 3.1.2 und 3.1.4 zeigen deutlich, wie empfindlich die Verfolgung auf Störungen reagiert, d.h. korrekte Anpassungen werden in der Gütefunktion (noch) nicht deutlich genug gegenüber Fehlanpassungen abgegrenzt. Insbesondere der Vergrößerung des Fangbereichs im Bild kommt hier große Bedeutung zu. In Abbildung 3.6 sind jeweils die prädierten Modelle eingezeichnet. Betrachtet man z.B. das aktualisierte Modell von Versuch 7 zum Halbbildzeitpunkt 281, so weicht dieses kaum ab vom prädierten, siehe Abbildung 3.6 oben rechts. Vom Aktualisierungsschritt sollte man aber erwarten können, daß er eine Abweichung zwischen Modellbein und Bildbein, wie in der Abbildung dargestellt, korrigieren kann. Dies ist ein besonders wichtiger Punkt. Eine Vergrößerung des Fangbereichs dürfte sich auch in vielen anderen Problemfällen stabilisierend auswirken. Beispielsweise geht dem Verlust eines Personenteils eine ungenaue Anpassung voraus.

Insgesamt lassen sich im wesentlichen folgende Hauptklassen verbleibender Anpassungsschwierigkeiten ausmachen.

- Die erste Klasse umfaßt Phänomene wie das im vorausgehenden Paragraphen angesprochene, bei denen ungenaue Anpassungen direkt auf einen zu kleinen Fangbereich im Bild durch den Aktualisierungsschritt zurückgeführt werden können (vgl. Abb. 3.6 oben rechts).
- Außerdem treten gelegentlich Effekte auf wie der in Abbildung 3.8 dargestellte, bei denen der Aperturereffekt die korrekte Schätzung von Längsbewegungen bestimmter Objektteile verhindert.
- Viele Anpassungsschwierigkeiten entstehen – oft nachdem ungenaue Anpassungen vorausgegangen sind – durch Anpassungen von Teilen des Personenmodells an den Hintergrund oder an falsche Personenteile. Insbesondere stellt die gegenseitige partielle Verdeckung der beiden Beine während des Laufens große Anforderungen an den Verfolgungsprozeß, vgl. Abbildung 3.9.
- Zum Teil sind letztere Effekte gekoppelt mit einem niedrigen Kontrast im Bereich der entsprechenden Bildteile. Abbildung 3.10 zeigt hierzu ein Beispiel.

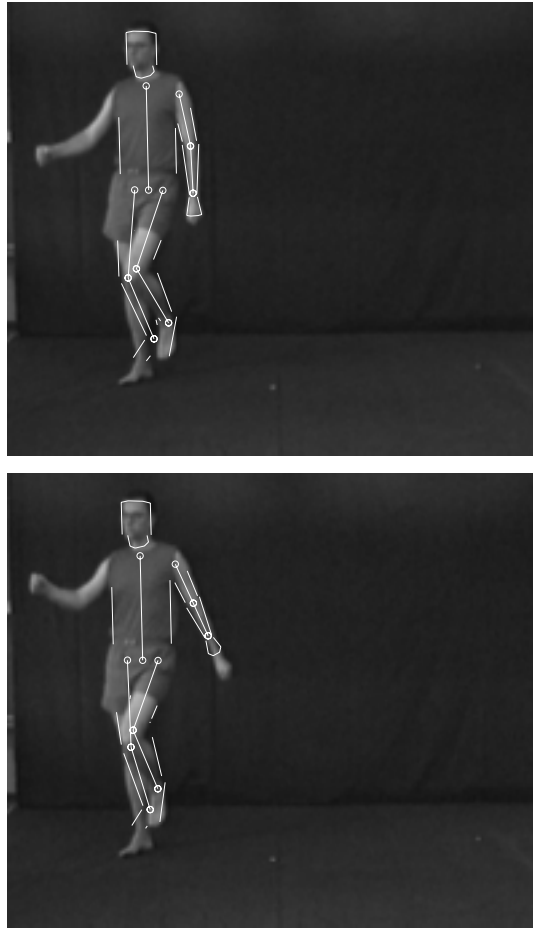


Abbildung 3.8: Modell Anpassungen bei einer Verfolgung zu `s1_191096` zu den Halbbildzeitpunkten 328 (oben) und 333 (unten). Die Bewegung des linken Armes wird aufgrund des Apertureffekts falsch eingeschätzt. Dargestellt sind die tiefpaßgefilterten Halbbilder mit dem angepaßten Modell.

3.2 Automatisierbarkeit

Unter dem Begriff der *Automatisierbarkeit* sei verstanden, die eingesetzten Programme möglichst viel selbstständig, d.h. ohne Eingriffe des Programmbenutzers, tun lassen zu können. Im Bezug auf die Personenverfolgung bedeutet dies als Fernziel, daß das Programm im Optimalfall eine neu eingespeiste Bildfolge ohne jegliche Benutzerinteraktion auszuwerten in der Lage ist. Manuelle Eingriffe sind also soweit wie möglich durch Automatismen zu ersetzen. Diesbezüglich gravierendster Punkt ist bei dem verwendeten Verfolgungssystem die bislang verwendete bildfolgenabhängige Parametrisierung des Verfolgungsalgorithmus. Zur Verbesserung der Automatisierung und damit der Selbständigkeit des Verfahrens ist daher eine Vereinheitlichung der Parameter für alle Bildfolgen vorzunehmen. Die Alternative dazu, die zu verwendenden Verfolgungs-

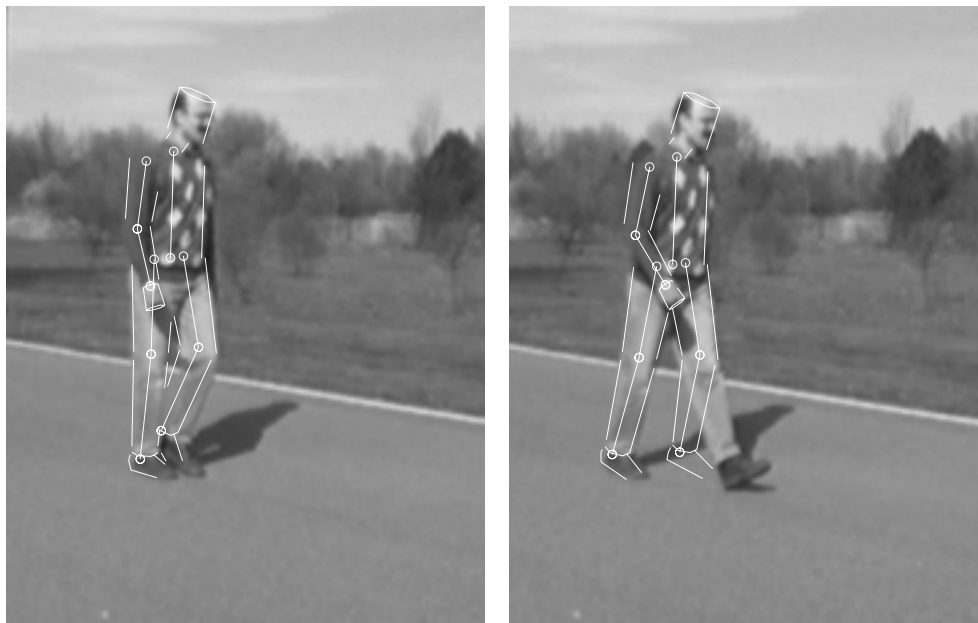


Abbildung 3.9: Modellanpassungen zu **frankm** zu den Halbbildzeitpunkten 96 (links) und 104 (rechts). Das linke Modellbein wird dem rechten mit angepaßt. Dargestellt sind die tiefpaßgefilterten Bilder mit dem jeweils angepaßten Modell.



Abbildung 3.10: Modellanpassung bei einer Verfolgung zu **hof_191096** zum Halbbildzeitpunkt 79. Es entsteht ein Verfolgungsproblem aufgrund des niedrigen Kontrasts zwischen den Beinen und dem Hintergrund. Dargestellt sind die tiefpaßgefilterten Halbbilder mit dem angepaßten Modell.

parameter aus dem Bildmaterial zu schätzen, wird nur in Erwägung gezogen, falls bei der Vereinheitlichung der Parameter unüberwindliche Schwierigkeiten auftreten sollten.

3.2.1 Übergang von einer bildfolgenindividuellen zu einer bildfolgenunabhängigen Parametrisierung

Die Parameter zur Steuerung der Verfolgung wurden ursprünglich von S. Wachter eingeführt und von ihm auf bestimmte, empirisch ermittelte optimale Werte gesetzt, die seither nicht verändert wurden. Daher existiert seither für jede Bildfolge ein eigener Parametersatz, mit dem die Modellanpassung vorgenommen wird. Dadurch werden zwar die Verfolgungsergebnisse zu den einzelnen Bildfolgen erwartungsgemäß optimiert, jedoch wird die Auswertung der Versuchsergebnisse erheblich erschwert, da sich die Ergebnisse zu verschiedenen Bildfolgen nicht ohne Weiteres vergleichen lassen. Zudem mußte bisher für jede neu hinzukommende Bildfolge ein neuer Parametersatz manuell bestimmt werden. Aufgrund der Optimierung der Parameter durch S. Wachter wird von der Vereinheitlichung der Parameter erwartet, daß im Mittel schlechtere Verfolgungsergebnisse produziert werden als bisher. Bis zu einem bestimmten Grad muß das auch in Kauf genommen werden. Die im weiteren verfolgte Methodik bei der Vereinheitlichung der Parameter soll sicherstellen, daß nach der Vereinheitlichung der Parameter die Verfolgung so gut wie möglich funktioniert. Da oben festgestellt wurde, daß der Verfolgungsalgorithmus ohnehin noch deutliche Schwachstellen aufweist, sind eventuell neu entstehende Ungenauigkeiten bei der Verfolgung sehr hilfreich zur Lokalisation und Beseitigung der zugrundeliegenden Ursachen. Die durch die Parametervereinheitlichung erzielte bessere Vergleichbarkeit der Verfolgungsergebnisse unterstützt den Prozeß der Robustheitssteigerung zudem erheblich.

In Tabelle 3.8 sind die Parameter zusammengestellt, die im Zusammenhang mit der Kameramodellierung eingesetzt werden. Tabelle 3.9 zeigt die übrigen Parameter, mit welchen sich der Anpassungsprozeß beeinflussen läßt. Die Parameter für die Personenmodelle sind nicht aufgeführt, denn diese sind ohne Belang für die nachfolgenden Betrachtungen, da sie zu jeder zu verfolgenden Person (d.h. zu jeder Bildfolge) separat eingestellt werden *müssen*.³ Bei Betrachtung der Tabelle 3.9 kann man einige Auffälligkeiten erkennen:

- Bei der Nutzung der Kanteninformation kommen bei den Bildfolgen zwei unterschiedliche Verfahren zum Einsatz: das Modellgradientenverfahren und das erweiterte Modellgradientenverfahren (vgl. [Wachter 97], S. 54 und S. 64). Das in [Wachter 97] ab Seite 62 beschriebene modifizierte Modellgradientenverfahren wird bei keiner Bildfolge benutzt. Der Benutzung eines einheitlichen Bewertungs-

³Unterschiedliche Personen können z.B. unterschiedlich groß sein. Die Schätzung der Größe einer Person aus dem Bildmaterial soll aber nicht Gegenstand dieser Untersuchungen sein. Dieser Punkt bedarf erst der näheren Zuwendung, wenn die Verfolgung für die manuell passend gewählten Personenmodelle hinreichend robust ist.

	l_rot	alinde	frankm	g_braun	hof_191096	sl_191096
kx, u	2650	2650	2650	2650	2650	2650
ky, u	2880	2880	2900	2880	2900	2880
ox, u	150	157	200	162	197	260
oy, u	198	150	180	242	210	208
Brennweite, u	0,3	0,3	0,3	0,3	0,61	0,3

Tabelle 3.8: Bisher eingesetzte Projektionsparameter bei den Bildfolgen. Es sind nur die Parameter u zu kx, ky, ox, oy und Brennweite aufgeführt. Die sonst noch einstellbaren Parameter v, pu, pv und q zu kx, ky, ox, oy bzw. zu Brennweite sind alle auf 0 eingestellt und in der Tabelle nicht aufgeführt, vgl. [Wachter 97].

	l_rot	alinde	frankm	g_braun	hof_191096	sl_191096
Kanteninformation	MG	EMG	EMG	MG	EMG	MG
Kantenabstand	3	3	1	4	3	2
Flächenabstand	3	2	1	4	3	6
Zylinderenden kompl.	<i>nein</i>	<i>ja</i>	<i>ja</i>	<i>nein</i>	<i>nein</i>	<i>nein</i>
Informationsnutzung	K, F	K, F	K, F	K, F	K, F	K, F
Meßrauschen Kanten	25	12	12	50	12	25
Meßrauschen Flächen	10	20	10	10	20	50
Meßrauschen Kopplung	0,1	0,1	0,1	0,1	0,1	0,1
Modellgradient, Kontrastfaktor	40	40	40	40	40	40
Modellgradient, sigma	3	3	3	3	3	3
Modellgradient, rel.	0,01	0,01	0,01	0,01	0,01	0,01
Mindestgröße						
maximale Grauwertdifferenz für Flächenmessung	256	25	25	256	256	256

Tabelle 3.9: Bisher eingesetzte Verfolgungsparameter bei den Bildfolgen. MG steht für ‘Modellgradient’ und EMG für ‘erweiterter Modellgradient’. Bei Informationsnutzung kann Kanteninformation, Flächeninformation oder beides eingestellt werden.

verfahrens ist eine höhere Priorität zuzuordnen als der Angleichung der Berechnungsparameter.

- Zur Bildfolge `frankm` ist bei den Parametern `Kantenabstand` und `Flächenabstand` der Wert 1 eingestellt, welcher deutlich kleiner ist als bei den anderen Bildfolgen. Dies dürfte erklären, warum vor allem bei den Versuchen zur Bildfolge `frankm` häufig Anpassungsprobleme im Zusammenhang mit Beinverdeckungen auftreten. Um zu verhindern, daß Modellteile den falschen Bildteilen mit angepaßt werden, dürfte ein größerer Wert als 1 grundsätzlich sinnvoll sein.
- Zur Bildfolge `s1_191096` sind bei den Parametern `Flächenabstand` und `Meßrauschen Flächen` deutlich größere Werte eingestellt als bei den anderen Bildfolgen, d.h. Flächeninformation wird deutlich weniger genutzt als bei den anderen Bildfolgen. Der bei `g_braun` eingestellte Wert für `Meßrauschen (Kanten)` ist deutlich größer als bei den anderen Bildfolgen, was bedeutet, daß die Kanteninformation schwächer gewichtet wird als bei den anderen Bildfolgen. Gemäß Seite 67 in [Wachter 97] wurde daher bei `s1_191096` verstärkt auf eine gute Lokalisierung des angepaßten Modells und bei `g_braun` auf eine robuste Verfolgung Wert gelegt.
- Der Parameter `maximale Grauwertdifferenz für Flächenmessung` beträgt bei vier Bildfolgen 256 und bei zwei Bildfolgen 25. Dies erscheint merkwürdig. Hier könnte möglicherweise bei der Eingabe des letzteren Wertes die letzte Ziffer vergessen worden sein.

Die Bedeutung des letzten Punktes wird anhand des Programmquelltextes untersucht: Bei der Nutzung der Flächeninformation wird geprüft, ob die gebildete Grauwertdifferenz den Wert `maximale Grauwertdifferenz für Flächenmessung` überschreitet. Falls dies der Fall ist, wird das entsprechende Pixel ignoriert, d.h. von der Flächenmessung ausgeschlossen. Ist also als Parameterwert 256 eingestellt, so findet ein solcher Ausschluß eines Pixels nie statt, da alle Grauwerte zwischen 0 und 255 liegen und somit die gebildete Differenz stets kleiner als 256 ist. Dies deutet darauf hin, daß bei den Bildfolgen `alinde` und `frankm` der genannte Ausschluß von Pixeln gezielt eingeschaltet wurde und bei den übrigen Bildfolgen von S. Wachter als nicht notwendig eingestuft wurde.

Zusammenfassend soll im Weiteren in der folgenden Reihenfolge vorgegangen werden:

1. Finden des besseren Modellgradientenverfahrens zur Nutzung von Kanteninformation.
2. Angleichung und Optimierung der Verfolgungsparameter für die Bildfolgen. Die Kameraparameter aus Tabelle 3.8 sollen nicht verändert werden, da sie die Beschaffenheit der möglicherweise unterschiedlichen zur Bildaufnahme verwendeten Kameras widerspiegeln.

Bei der Durchführung ist natürlich eine manuelle erschöpfende Suche im gesamten Parameterraum aus Zeit- und Zweckmäßigkeitsgründen auszuschließen. Stattdessen wird eine Form eines ‘manuellen Gradientenabstiegs’ zur Vereinheitlichung der Parameter gewählt, bei dem als Gütefunktion die Genauigkeit der durchgeführten Modellanpassungen Verwendung findet und mögliche Sackgassen möglichst frühzeitig vermieden werden.

Finden des besseren Modellgradientenverfahrens

Das bei keiner Bildfolge benutzte modifizierte Modellgradientenverfahren soll auch im Weiteren nicht untersucht werden, da es offenbar von S. Wachter als schlechter eingestuft wurde als die anderen beiden Modellgradientenverfahren und so der Aufwand für das Finden des besten Verfahrens reduziert werden kann. Nun muß entschieden werden, ob das Modellgradientenverfahren oder das erweiterte Modellgradientenverfahren für den Anpassungsprozeß geeigneter ist. Um das bessere Verfahren herauszufinden, werden in den Versuchen 15 bis 20 die Versuche 1 bis 6 wiederholt unter Verwendung des jeweils bisher nicht benutzten Modellgradientenverfahrens. Ein Vergleich der Versuche 15 bis 20 mit den Versuchen 1 bis 6 soll zeigen, welches Modellgradientenverfahren zur Fußgängerverfolgung besser geeignet ist (vgl. Tabelle 3.10).

Die Versuche 15 bis 20 wurden durchgeführt, um das erweiterte Modellgradientenverfahren mit dem Modellgradientenverfahren zu vergleichen. Bereits bei den Versuchen 15 (Modellgradientenverfahren) und 16 (erweitertes Modellgradientenverfahren) ergeben sich deutlich schlechtere Ergebnisse als in den Versuchen 1 und 2. Auf den ersten Blick scheinen auch die Versuche 17 bis 20 schlechtere Ergebnisse zu liefern als die Versuche 3 bis 6. Dies deutet auf eine Korrelation des eingestellten Modellgradientenverfahrens mit einem oder mehreren einstellbaren Parametern hin. Sollte dies zutreffen, wird ein Vergleich der beiden Modellgradientenverfahren mittels der Versuche 15 bis 20 sehr schwierig ausfallen oder sogar unmöglich sein. Tatsächlich kann man in der Tabelle 3.9 erkennen, daß bei den Bildfolgen zum erweiterten Modellgradientenverfahren bei **Meßrauschen Kanten** der Wert 12 eingestellt ist, während sich hier bei den anderen Bildfolgen die deutlich größeren Werte 25 bzw. 50 finden. Daher werden die Auswertungen der Versuche 17 bis 20 als nicht zweckdienlich eingestuft und deshalb nicht durchgeführt. Zum Vergleich der beiden Modellgradientenverfahren sollen statt der Versuche 15 bis 20 die Versuche 21 bis 26 benutzt werden, bei denen die erkannte Korrelation berücksichtigt wird, siehe Tabelle 3.11. Eine weitere evtl. bestehende Korrelation zwischen dem verwendeten Modellgradientenverfahren und den Parametern ist aus Tabelle 3.9 nicht ohne Weiteres zu erkennen. Es wird erwartet, daß sich beim Vergleich der Versuche 21 bis 26 mit den Versuchen 1 bis 6 das erweiterte Modellgradientenverfahren als das bessere Verfahren herausstellt, da nicht nur der Betrag des Grauwertgradienten berücksichtigt wird, sondern durch die Betrachtung der beiden Komponenten des Grauwertgradienten auch Richtungsinformation genutzt wird.

Sollte sich auch bei Verwendung der Versuche 21 bis 26 nur schwer feststellen lassen,

Vers.	Bildfolge	Bemerkungen
15	<code>l_rot</code>	Wiederholung von Versuch 1 unter Verwendung des erweiterten Modellgradientenverfahrens. Im Vergleich zu Versuch 1 ergeben sich zwischen 42 und 67 schlechtere Anpassungen des rechten Armes und zwischen 103 und 112 schlechtere Anpassungen des linken Beines, als dieses einen schwachen Kontrast zum Hintergrund aufweist. Ab 393 treten schlechtere Anpassungen der bereits in Versuch 1 schlecht angepaßten Beine auf.
16	<code>alinde</code>	Wiederholung von Versuch 2 unter Verwendung des Modellgradientenverfahrens. Zwischen 29 und 357 erfolgen im Vergleich zu Versuch 2 wesentlich schlechtere Anpassungen des Kopfes, zwischen 16 und 35 bessere Anpassungen des linken Oberarmes, zwischen 58 und 96 sowie zwischen 245 und 357 schlechtere Anpassungen des linken Oberarmes, zwischen 230 und 245 schlechtere Rumpfanpassungen, ab 299 schlechtere Anpassungen des rechten Beins, und zwischen 326 und 348 erfolgen bessere Anpassungen des linken Unterschenkels.
17	<code>frankm</code>	Wiederholung von Versuch 3 unter Verwendung des Modellgradientenverfahrens. ...
18	<code>g_braun</code>	Wiederholung von Versuch 4 unter Verwendung des erweiterten Modellgradientenverfahrens. ...
19	<code>hof_191096</code>	Wiederholung von Versuch 5 unter Verwendung des Modellgradientenverfahrens. ...
20	<code>s1_191096</code>	Wiederholung von Versuch 6 unter Verwendung des erweiterten Modellgradientenverfahrens. ...

Tabelle 3.10: Versuche 15 bis 20 zur Angleichung des Modellgradientenverfahrens bei den Bildfolgen. Unterschiede zu den Referenzversuchen: Zwangssymmetrisierungen, Minimierungsverfahren, Modellgradientenverfahren. Die drei Punkte bei den Versuchen 17 bis 20 deuten jeweils darauf hin, daß sich die Ergebnisauswertungen nach Durchführung der Versuche im Hinblick auf das weitere Vorgehen als nicht wichtig herausgestellt haben, siehe Text.

welches Modellgradientenverfahren das bessere ist, so kann nötigenfalls auch anders vorgegangen werden: Jede Bildfolge behält zunächst ihr ursprüngliches Modellgradientenverfahren. Getrennt nach dem Modellgradientenverfahren können anschließend alle einstellbaren Parameter vereinheitlicht werden, um einen einheitlichen, optimalen Parametersatz zu jedem der beiden Verfahren zu erhalten. Anschließend kann bei jeder der Bildfolgen das jeweils andere Modellgradientenverfahren samt den so gefundenen

Vers.	Bildfolge	Bemerkungen
21	l_rot	Wiederholung von Versuch 1 unter Verwendung des erweiterten Modellgradientenverfahrens. Im Vergleich zu Versuch 1 ergeben sich im Halbbildzeitraum 32 bis 69 wesentlich schlechtere Anpassungen des rechten Armes, im Zeitraum 331 bis 365 bessere Anpassungen des rechten Armes, zwischen 414 und 455 bessere Anpassungen des linken Unterschenkels, zwischen 482 und 518 bessere Anpassungen des linken Unterschenkels bzw. des ganzen Beins und im Zeitraum 492 bis 525 bessere Rumpfanpassungen.
22	alinde	Wiederholung von Versuch 2 unter Verwendung des Modellgradientenverfahrens. Gegenüber Versuch 2 erfolgen zwischen 37 und dem Bildfolgenende schlechtere Kopfanpassungen, zwischen 10 und 51 bessere Anpassungen des linken Oberarmes, im Zeitraum 224 bis 226 schlechtere Anpassungen des rechten Unterschenkels (Anpassung an verdeckendes, linkes Bein), zwischen 218 und 229 schlechtere Anpassungen des linken Oberarmes, zwischen 249 und 357 bessere Anpassungen des rechten Beins, und im Zeitraum 274 bis 293 treten bessere Rumpfanpassungen auf (Rumpf wird nicht so stark dem Hintergrund mit angepaßt wie in Versuch 2).
23	frankm	Wiederholung von Versuch 3 unter Verwendung des Modellgradientenverfahrens. Gegenüber Versuch 3 ergeben sich zwischen 69 und 149 im Zusammenhang mit einer Beinverdeckung schlechtere Anpassungen des linken Beins, ab 448 schlechtere Anpassungen des rechten Armes und ab Halbbild 410 schlechtere Personenanpassungen.
24	g_braun	Wiederholung von Versuch 4 unter Verwendung des erweiterten Modellgradientenverfahrens. Im Vergleich zu Versuch 4 erfolgen zwischen 87 und 109 sowie zwischen 357 und 407 schlechtere Anpassungen des linken Unterarmes und im Zeitraum 114 bis 124 bessere Anpassungen des linken Unterarmes.
25	hof_191096	Wiederholung von Versuch 5 unter Verwendung des Modellgradientenverfahrens. Ab Halbbild 40 werden schlechtere Anpassungen des rechten Unterschenkels bzw. Beins beobachtet als in Versuch 5.
26	s1_191096	Wiederholung von Versuch 6 unter Verwendung des erweiterten Modellgradientenverfahrens. Jeweils im Zusammenhang mit einer Beinverdeckung ergeben sich zwischen 202

Vers.	Bildfolge	Bemerkungen
		und 210 bessere Anpassungen des rechten Beins und ab Halbbild 261 schlechtere Anpassungen des rechten Beins als in Versuch 6.

Tabelle 3.11: Versuche 21 bis 26 zur Angleichung des Modellgradientenverfahrens bei den Bildfolgen. Bei Verwendung des erweiterten Modellgradientenverfahrens wird jeweils für **Meßrauschen Kanten** der Wert 12 benutzt und bei Verwendung des Modellgradientenverfahrens der Wert 25. Unterschiede zu den Referenzversuchen: Zwangssymmetrisierungen, Minimierungsverfahren, Modellgradientenverfahren, **Meßrauschen Kanten**.

Parametern erprobt werden.

Die Auswertung der Versuche 21 bis 26 führt zu folgendem Ergebnis: In insgesamt 13 Situationen ergeben sich bei Verwendung des erweiterten Modellgradientenverfahrens bessere Anpassungen und in nur 7 Situationen schlechtere Anpassungen als bei Verwendung des Modellgradientenverfahrens. Dies zeigt eindeutig, daß das erweiterte Modellgradientenverfahren zur Fußgängerfolgung besser geeignet ist als das Modellgradientenverfahren, was den eingangs geäußerten Erwartungen entspricht. Von nun an wird daher ausschließlich mit dem erweiterten Modellgradientenverfahren gearbeitet.

Angleichen der Parameter Kantenabstand und Flächenabstand

Als nächstes sollen die Parameter **Kantenabstand** und **Flächenabstand** für alle Bildfolgen einheitlich gewählt werden. Diese Parameter sollen als erstes angeglichen werden, da einige bisher beobachtete Verfolgungsprobleme mit einer schlechten Einstellung dieser Parameter direkt erklärbar sind (Versuche zu **frankm** und **s1_191096**). Die beiden Parameter werden als korreliert betrachtet und müssen daher abhängig voneinander angeglichen werden. Zur Einschränkung des Suchraumes für die beiden Parameter sollen **Kantenabstand** und **Flächenabstand** gleich gewählt werden. Zum einen erscheint dies nach [Wachter 97] (Seiten 49 und 50) plausibel und zum anderen wurden die beiden Parameter bisher in vier der sechs Bildfolgen gleich gewählt, vgl. Tabelle 3.9. In Tabelle 3.9 kann man auch erkennen, daß am häufigsten der Wert 3 für die Parameter benutzt wurde. Dies dürfte ein guter Wert für die Parameter sein. Aus diesem Grund wird als Ausgangspunkt **Kantenabstand** = **Flächenabstand** = 3 gewählt. Die durchgeführten Versuche 27 bis 30 sind in Tabelle 3.12 den entsprechenden Ergebnissen der Versuche 2, 3, 24 und 26 gegenübergestellt (einziger Unterschied: Parameter **Kantenabstand** und **Flächenabstand**). Die Versuche zu **l_rot** und **hof_191096** wurden bereits in den Versuchen 21 und 5 durchgeführt und tauchen daher in der Tabelle nicht auf.

Betrachtet man die wesentlichen Unterschiede zwischen den Versuchen 27 bis 30 und den Versuchen 2, 3, 24 und 26 (Tabelle 3.12), so fällt folgendes auf:

- Zwischen den Versuchen zu **g_braun** ergeben sich keine Unterschiede, obwohl

Vers.	Bildfolge	Bemerkungen
27	alinde	Gegenüber Versuch 2 erfolgen im Zusammenhang mit einer Verdeckung durch das linke Bein zwischen 196 und 237 schlechtere Anpassungen des rechten Unterschenkels und zwischen 246 und 268 bessere Anpassungen des rechten Beins. Zwischen 245 und 267 stellt man schlechtere Anpassungen des linken Oberarmes und im Zeitraum 331 bis 346 bessere Anpassungen des linken Unterschenkels fest als in Versuch 2.
28	frankm	Im Vergleich zu Versuch 3 ergeben sich jeweils im Zusammenhang mit einer Verdeckung durch das rechte Bein zwischen 94 und 127 sowie zwischen 153 und 178 schlechtere Anpassungen des linken Unterschenkels. Im Zeitraum 146 bis 151 werden bessere Anpassungen des linken Unterschenkels und zwischen 323 und 351 schlechtere Anpassungen des rechten Armes festgestellt.
29	g_braun	Es werden keine wesentlichen Unterschiede zu Versuch 24 beobachtet.
30	s1_191096	Im Vergleich zu Versuch 26 treten ab 264 bessere Anpassungen des rechten Beins auf (Modellbein wird nicht dem linken, verdeckenden Bildbein mit angepaßt wie in Versuch 26). Zwischen 355 und 378 ergeben sich bessere Anpassungen des linken Armes als in Versuch 26.

Tabelle 3.12: Versuche 27 bis 30 zur Angleichung der Parameter **Kantenabstand** und **Flächenabstand**. Es wird jeweils das erweiterte Modellgradientenverfahren benutzt bei **Meßrauschen** $Kanten = 12$ und $Kantenabstand = Flächenabstand = 3$. Die übrigen Parameter bleiben unverändert und können Tabelle 3.9 entnommen werden. Die Versuche zu **l_rot** und **hof_191096** entsprechen den Versuchen 87 und 79. Unterschiede zu den Referenzversuchen: Zwangssymmetrisierungen, Minimierungsverfahren, Modellgradientenverfahren, o.g. Parameter.

einige Unterschiede erwartet wurden aufgrund des Parameterunterschiedes in **Kantenabstand** und **Flächenabstand** zwischen den Versuchen. Dagegen erscheint die Anzahl an Unterschieden zwischen den Versuchen zu **alinde** hoch, schließlich sind die Parameterunterschiede hier geringer als bei den Versuchen zu **g_braun**.

- Durch die Angleichung $Kantenabstand = Flächenabstand = 3$ bleiben die Versuchsergebnisse zu **alinde** in etwa so gut wie zuvor.
- Bei **frankm** verschlechtern sich die Ergebnisse, wenn für **Kantenabstand** und **Flächenabstand** anstatt des Wertes 1 der Wert 3 benutzt wird. Dies verläuft

entgegen den Erwartungen, welche auf Seite 66 geäußert wurden.

- Bei `s1_191096` ergibt sich in Versuch 30 gegenüber Versuch 26 eine deutliche Verbesserung. Offenbar wurden die Parameter von S. Wachter doch nicht so weit für jede Bildfolge optimiert, wie anfangs anzunehmen war.
- Insgesamt treten in den Versuchen 27 bis 30 gegenüber den Versuchen 2, 3, 24 und 26 5 wesentliche Verbesserungen und genau so viele Verschlechterungen auf. Das bedeutet, daß sich die Ergebnisse durch die Angleichung **Kantenabstand** = **Flächenabstand** = 3 im Mittel nicht verschlechtert haben. Dies ist erfreulich, nachdem mit Verschlechterungen zu rechnen war.

Der letzte Punkt zeigt, daß die Versuche 5, 21 und 27 bis 30 eine gute Ausgangsbasis bilden für weitere Parameteroptimierungen und -angleichungen. Die Punkte drei und vier zeigen, daß sich weitere Untersuchungen zu den Parametern **Kantenabstand** und **Flächenabstand** lohnen, zum einen, weil deren Auswirkung auf den Verfolgungsprozeß offenbar noch nicht genau genug verstanden ist, und zum anderen, weil eine mögliche Verbesserung des Anpassungsprozesses in Aussicht steht. Die Punkte eins und drei lassen vermuten, daß mittels der Wahl **Kantenabstand** = **Flächenabstand** = 2 möglicherweise eine Verbesserung erreicht werden kann gegenüber den Versuchen mit **Kantenabstand** = **Flächenabstand** = 3. Daher werden in den Versuchen 31 bis 36 die Versuche 21, 27, 28, 29, 5 und 30 wiederholt unter Verwendung des Wertes 2 für die beiden Parameter, siehe Tabelle 3.13.

Zusammenfassend kann folgendes gesagt werden: Bei den Versuchen zu `1_rot` und `alinde` ergibt sich eine eindeutige Verbesserung, wenn **Kantenabstand** = **Flächenabstand** = 2 anstatt **Kantenabstand** = **Flächenabstand** = 3 benutzt wird, bei den Versuchen zur Bildfolge `frankm` wird eine Verschlechterung festgestellt und bei den restlichen Bildfolgen ändern sich die Verfolgungsergebnisse bei der genannten Änderung der Parameter (im Mittel) nicht. Insgesamt treten 6 Situationen auf, bei denen sich eine Verbesserung ergibt und 4 Situationen, bei denen eine Verschlechterung der Anpassung eintritt. Das bedeutet, daß sich mit **Kantenabstand** = **Flächenabstand** = 2 bessere Ergebnisse erzielen lassen als mit **Kantenabstand** = **Flächenabstand** = 3, weshalb im folgenden der Wert 2 für die beiden Parameter benutzt werden soll.

Verbesserung des Personenmodells für die Bildfolge `frankm`

Die Tatsache, daß sich in den Versuchen 31 und 32 eindeutig bessere Resultate ergeben als in den Versuchen 21 und 27, die Ergebnisse der Versuche 34 bis 36 ähnlich gut sind wie die der Versuche 29, 5 bzw. 30 und nur in Versuch 33 im Vergleich zu Versuch 28 eindeutig schlechtere Ergebnisse beobachtet werden, deutet auf einen besonderen Störfaktor bei der Bildfolge `frankm`. Auffällig ist auch, daß sich bei dieser Bildfolge bei Verwendung des Wertes 2 für die Parameter **Kantenabstand** und **Flächenabstand** schlechtere Ergebnisse ergeben als bei Verwendung der Werte 1 oder 3 (Versuche 3, 28 und 33). Die noch nicht angeglichenen Parameter und die Projektionsparameter weisen

Vers.	Bildfolge	Bemerkungen
31	l_rot	Wiederholung von Versuch 21. Im Vergleich zu Versuch 21 treten zwischen 32 und 67 bessere Anpassungen des rechten Armes auf und im Zeitraum 399 bis 561 bessere Anpassungen der rechten Schulter.
32	alinde	Wiederholung von Versuch 27. Gegenüber Versuch 27 erfolgen zwischen 197 und 241 im Zusammenhang mit einer Verdeckung durch das linke Bein bessere Anpassungen des rechten Unterschenkels, im Zeitraum 212 bis 246 erfolgen bessere Anpassungen des Rumpfes und des Beckens, zwischen 242 und 270 bessere Anpassungen des linken Oberarmes und ab 246 schlechtere Anpassungen des rechten Beins im Zusammenhang mit einer Verdeckung durch das linke Bein.
33	frankm	Wiederholung von Versuch 28. Jeweils im Zusammenhang mit einer Verdeckung durch das rechte Bein erfolgen zwischen 154 und 185 schlechtere Anpassungen des linken Unterschenkels und zwischen 186 und 226 schlechtere Anpassungen des linken Beins als in Versuch 28.
34	g_braun	Wiederholung von Versuch 29. Es treten keine wesentlichen Unterschiede zu den Anpassungen von Versuch 29 auf.
35	hof_191096	Wiederholung von Versuch 5. Im Vergleich zu Versuch 5 treten zwischen 9 und 27 bessere Anpassungen des linken Unterschenkels auf, als die Person noch steht. Zwischen 75 und 114 ergeben sich schlechtere Anpassungen des rechten Unterschenkels im Zusammenhang mit einer Verdeckung durch das linke Bein und schwachem Kontrast zum Hintergrund.
36	s1_191096	Wiederholung von Versuch 30. Es treten keine erwähnenswerten Unterschiede zu Versuch 30 auf.

Tabelle 3.13: Versuche 31 bis 36: Wiederholung der Versuche 21, 27, 28, 29, 5 und 30 unter Verwendung von $\text{Kantenabstand} = \text{Flächenabstand} = 2$ zur Parameteroptimierung. Unterschiede zu den Referenzversuchen: Zwangssymmetrisierungen, Minimierungsverfahren, erweitertes Modellgradientenverfahren, $\text{Meßrauschen Kanten} = 12$, $\text{Kantenabstand} = \text{Flächenabstand} = 2$.

bei **frankm** keine Besonderheiten auf (vgl. Tabellen 3.8 und 3.9). Das Problem rührt offenbar daher, daß die zu verfolgende Person bei der Bildfolge **frankm** eine sehr weit geschnittene Hose trägt, durch welche die Bildunterschenkel oft deutlich breiter werden als die Modellunterschenkel. In Abbildung 3.11 sind zwei Situationen aus Versuch 33 gezeigt, an denen man dies gut nachprüfen kann. Es liegt nahe, viele der bereits im

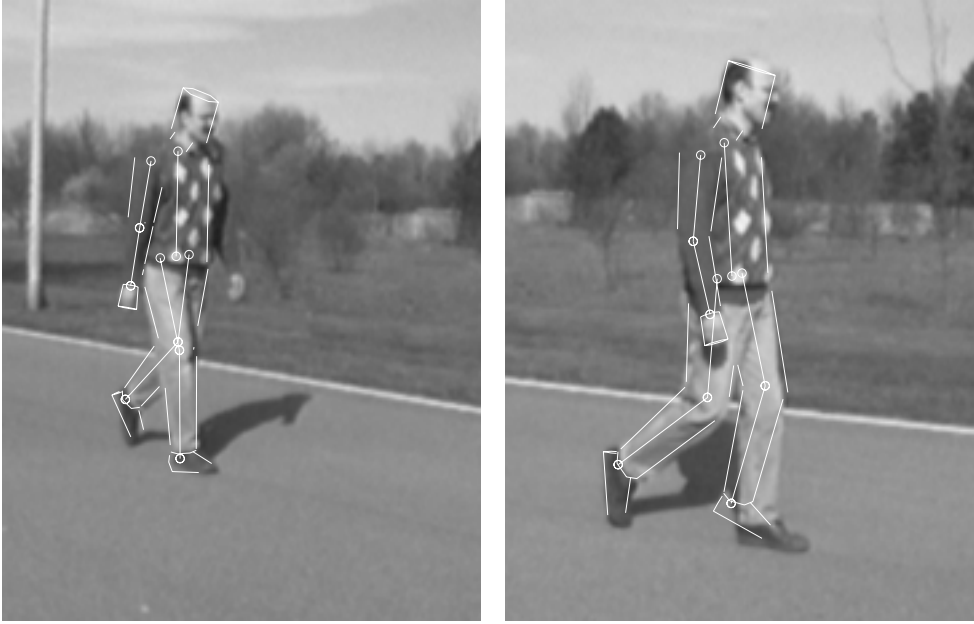


Abbildung 3.11: Modellanpassungen von Versuch 33 zu den Halbbildzeitpunkten 86 (links) und 176 (rechts). Die Modellunterschenkel haben einen deutlich kleineren Durchmesser als die Bildunterschenkel. Dargestellt sind die tiefpaßgefilterten Halbbilder mit dem angepaßten Modell.

Zusammenhang mit Beinverdeckungen beobachteten Verfolgungsprobleme bei **frankm** auf die schlecht gewählte Größe der Modellunterschenkel zurückzuführen bzw. auf die sich im Verlauf der Bildfolge ändernde Breite der Bildunterschenkel. Es soll versucht werden, das Problem zu beheben. Dies wird an dieser Stelle als wichtig erachtet, denn bereits bei den Versuchen 31 bis 36 hat sich **frankm** als ‘Ausreißer’ entpuppt. Dies könnte in nachfolgenden Versuchen eine Störung der Versuchsauswertungen bedeuten, wenn der Versuch zu **frankm** nicht vernachlässigt werden soll. Die Alternative, die Bildfolge **frankm** nicht mehr zu verwenden oder die Störungen beizubehalten, kommt nicht in Frage, denn die Verwendung jeder zusätzlichen Bildfolge macht den sich weiter entwickelnden Verfolgungsprozeß allgemeiner einsetzbar.

Um eine bessere Unterschenkelanpassung zu ermöglichen und dadurch das Problem zu beheben, sollen die Modellunterschenkel verbreitert werden. In Versuch 33 sind die Radien der Modellunterschenkel am Knieende $9,5\text{ cm}$ und am Fußende 7 cm . In Versuch 37 wird Versuch 33 wiederholt unter Verwendung der Radien $7,7\text{ cm}$ am Fußende, in Versuch 38 werden die Radien $8,7\text{ cm}$ benutzt (Tabelle 3.14). Obwohl die Breite der Bildunterschenkel im Verlauf der Bildfolge **frankm** variiert, wird erwartet, daß einer der Versuche 37 bzw. 38 bessere Ergebnisse liefert als Versuch 33.

Die in Versuch 38 gegenüber Versuch 33 auftretenden besseren Anpassungen der Unterschenkel entsprechen den Erwartungen. Das Modell mit den breiteren Unterschenkeln kann besser an die Bilddaten angepaßt werden, wodurch sich einige Fehlanpassungen

Vers.	Bildfolge	Bemerkungen
37	frankm	Radien des Modellunterschenkels am Fußende: 7,7 cm. Im Vergleich zu Versuch 33 ergeben sich zwischen 127 und 143 schlechtere Anpassungen des linken Unterschenkels im Zusammenhang mit einer Verdeckung durch das rechte Bein, zwischen 147 und 203 wesentlich bessere Anpassungen des linken Unterschenkels und zwischen 207 und 231 bessere Anpassungen des linken Oberschenkels.
38	frankm	Radien des Modellunterschenkels am Fußende: 8,7 cm. Im Vergleich zu Versuch 33 erfolgen zwischen 88 und 122 sowie zwischen 149 und 185 wesentlich bessere Anpassungen des linken Unterschenkels jeweils im Zusammenhang mit einer Verdeckung durch das rechte Bein, zwischen 175 und 296 erfolgen bessere Anpassungen des linken Oberschenkels, ab 323 treten schlechtere Anpassungen der rechten Hand und dadurch ab 415 schlechtere Anpassungen des rechten Unterarmes auf.

Tabelle 3.14: Versuche 37 und 38 Wiederholung von Versuch 33 unter Verwendung breiterer Modellunterschenkel.

des linken Unterschenkels nach auftretenden Verdeckungen durch das rechte Bein vermeiden lassen. Zur Illustration zeigt Abbildung 3.12 die Ergebnisse der Versuche 33 und 38 zum Halbbildzeitpunkt 161. Bis Halbbild 156 tritt eine Verdeckung des linken Beins durch das rechte auf. Man kann in der Abbildung erkennen, wie anschließend in Versuch 38 eine bessere Anpassung des linken Unterschenkels erfolgt und daß die breiteren Modellunterschenkel gut zu den Bildunterschenkeln passen. Die Situationen, in welchen die Bildunterschenkel durch die anliegenden Hosenbeine besonders schmal werden, bereiten offenbar in Versuch 38 keine Schwierigkeiten, siehe hierzu Abbildung 3.13.

Die unterschiedliche Handanpassung bei den Versuchen 33 und 38 erfolgt unerwartet, da sich die Versuche 33 und 38 ausschließlich in der Breite der Modellunterschenkel unterscheiden und eine Beeinflussung der Handanpassung durch den Parameter für die Modellunterschenkeldicke nicht erwartet wurde. Die Unterschiede bei der Anpassung der rechten Hand entstehen nicht abrupt. Bei genauerer Betrachtung der Versuche 33 und 38 erkennt man, daß sich zunächst kleine Unterschiede langsam aufschaukeln, bis sie dem Betrachter deutlich sichtbar werden. Ein solches Verhalten ist bei genauerer Überlegung nicht verwunderlich, schließlich kann eine Korrelation zwischen Unterschenkel- und Armanpassungen leicht mittels der Invertierungen der Kovarianzmatrizen P_k^- und P_k^+ erklärt werden: Bei einer von einem Referenzversuch abweichenden Anpassung des Unterschenkels verändert sich die approximierete Hessematrix $\overline{\nabla^2\Gamma}$ der Gütefunktion Γ und damit auch die Kovarianzmatrix $P_k^+ = (\overline{\nabla^2\Gamma})^{-1}$ (vgl. S. 47),



Abbildung 3.12: Modellanpassungen von Versuch 33 (links) und Versuch 38 (rechts) zum Halbbildzeitpunkt 161. Dargestellt sind die tiefpaßgefilterten Halbbilder mit dem angepaßten Modell.



Abbildung 3.13: Modellanpassungen von Versuch 33 (links) und Versuch 38 (rechts) zum Halbbildzeitpunkt 53. Dargestellt sind die tiefpaßgefilterten Halbbilder mit dem angepaßten Modell.

welche zur Bestimmung der Kovarianzmatrix P_{k+1}^- dient und von deren Inversen wiederum die Gütefunktion Γ des folgenden Aktualisierungsschrittes abhängt (vgl. Gleichungen (3.4), (3.41) und (3.52) in [Wachter 97]). Bei den beiden Inversenberechnungen kann eine Veränderung der zu invertierenden Matrix in den Einträgen, welche zu den Unterschenkelanpassungen gehören, zu einer (wenn auch kleinen) Veränderung der Inversen in den Einträgen führen, die den Handanpassungen zugeordnet werden. Dies kann damit begründet werden, daß ein Verändern der ‘Unterschenkeleinträge’ der Ausgangsmatrix i.a. bereits deren Determinante verändert, was sich auf alle Einträge der Inversen niederschlägt ($A^{-1} = A^\# / \det(A)$, $A^\#$ ist die Adjunkte der Matrix A). Aus den bisherigen Untersuchungen ist bekannt, wie entstehende kleine Veränderungen sich aufschaukeln können.

Angleichung und Optimierung des Parameters Meßrauschen Flächen

Die Tatsache, daß die Versuche 31 bis 36 (im Mittel) bessere Ergebnisse liefern als die Versuche 21, 27, 28, 29, 5 und 30 könnte bedeuten, daß sich für **Kantenabstand** = **Flächenabstand** = 1 noch bessere Ergebnisse ergeben. Auch die Versuche 3, 28 und 33 zu **frankm** könnten darauf hinweisen. Da eine Korrelation zwischen den Parametern **Flächenabstand** und **Meßrauschen Flächen** besteht, soll zuerst der Parameter **Meßrauschen Flächen** für alle Bildfolgen vereinheitlicht und optimiert werden. Anschließend soll untersucht werden, ob die Verwendung des Wertes 1 für die Parameter **Kantenabstand** und **Flächenabstand** eine weitere Verbesserung der Verfolgungsergebnisse bringt.

Zwischen den Parametern **Meßrauschen Flächen** und **Meßrauschen Kanten** besteht eine direkte Korrelation, d.h. die beiden Parameter müßten eigentlich zusammen optimiert werden. Der Parameter **Meßrauschen Kanten** wurde bereits vereinheitlicht. Um den Suchraum zu begrenzen, soll dieser Parameter nicht mehr verändert werden, d.h. lediglich der Parameter **Meßrauschen Flächen** ist zu vereinheitlichen und zu optimieren. Bisher wurden für **Meßrauschen Flächen** die Werte 10, 20 und 50 verwendet. Letzterer Wert wird als weniger geeignet eingestuft als die ersten beiden Werte, auch deshalb, weil er nur in einer Bildfolge verwendet wurde. Daher soll unter Verwendung der Werte 10 und 20 mittels der Versuche 39 bis 45 entschieden werden, welcher der beiden Werte für die Fußgängerfolgung besser geeignet ist (vgl. Tabelle 3.15).

Durch Auswertung der Versuche 39 bis 45 gelangt man zu folgendem Ergebnis: Der Verfolgungsprozeß wird durch die Verwendung des Parameterwertes 20 für **Meßrauschen Flächen** gegenüber dem Wert 10 bei vier Bildfolgen schlechter, bei einer etwas besser (**s1_191096**) und bei einer bleibt das Ergebnis im Mittel gleich (**g_braun**). Insgesamt ergibt sich bei Verwendung des Wertes 20 in 9 Situationen eine Verbesserung und in 15 Situationen eine Verschlechterung gegenüber **Meßrauschen Flächen** = 10. Der Wert 10 ist also eindeutig dem Wert 20 für **Meßrauschen Flächen** vorzuziehen. Die hohe Anzahl an Verbesserungen und auch an Verschlechterungen zwischen den Versuchen mit **Meßrauschen Flächen** = 10 und denen mit **Meßrauschen Flächen** = 20 sowie die Tatsache, daß die getesteten Parameterwerte relativ weit auseinander liegen (der eine

Vers.	Bildfolge	Bemerkungen
39	l_rot	Wiederholung von Versuch 31 bei Meßrauschen Flächen = 20. Im Zusammenhang mit einer Verdeckung durch das rechte Bein erfolgen zwischen 366 und 375 sowie zwischen 396 und 425 schlechtere Anpassungen des linken Unterschenkels als in Versuch 31. Zwischen 371 und 387 treten bessere Anpassungen des linken Knies auf als in Versuch 31.
40	alinde	Wiederholung von Versuch 32 bei Meßrauschen Flächen = 10. Jeweils im Zusammenhang mit einer Verdeckung durch das linke Bein treten im Vergleich zu Versuch 32 zwischen 195 und 245 schlechtere Anpassungen des rechten Unterschenkels und ab 245 bessere Anpassungen des rechten Oberschenkels auf. Zwischen 270 und 292 erfolgen bessere Rumpfanpassungen als in Versuch 32 (schwächere Anpassung an den Hintergrund als in Versuch 32)
41	frankm	Wiederholung von Versuch 38 bei Meßrauschen Flächen = 20. Im Vergleich zu Versuch 38 ergeben sich zwischen 94 und 118 wesentlich schlechtere Anpassungen des linken Unterschenkels, zwischen 324 und 349 schlechtere Anpassungen des rechten Ellenbogens, ab 325 bessere Anpassungen der rechten Hand und dadurch ab 415 bessere Anpassungen des rechten Unterarmes.
42	g_braun	Wiederholung von Versuch 34 bei Meßrauschen Flächen = 20. Gegenüber Versuch 34 erfolgen zwischen 89 und 109 schlechtere Anpassungen des linken Unterarmes und im Zeitraum 326 bis 356 bessere Anpassungen des linken Unterarmes.
43	hof_191096	Wiederholung von Versuch 35 bei Meßrauschen Flächen = 10. Im Vergleich zu Versuch 35 beobachtet man zwischen 8 und 24 schlechtere Anpassungen des linken Unterschenkels, ab 75 wesentlich bessere Anpassungen des rechten Unterschenkels (Verdeckung durch linkes Bein, schwacher Kontrast zum Hintergrund), zwischen 100 und 132 wesentlich bessere Anpassungen des Beckens, im Zusammenhang mit einer Verdeckung durch das linke Bein zwischen 136 und 150 bessere Anpassungen des rechten Oberschenkels, zwischen 157 und 166 bessere Anpassungen des rechten Knies und zwischen 233 und 257 bessere Anpassungen von Rumpf und Becken.
44	s1_191096	Wiederholung von Versuch 36 bei Meßrauschen Flächen = 20. Zwischen 292 und 313 ergeben sich schlechtere Anpas-

Vers.	Bildfolge	Bemerkungen
45	s1_191096	sungen des rechten Unterschenkels als in Versuch 36. Wiederholung von Versuch 36 bei Meßbrauschen Flächen = 10. Im Vergleich zu Versuch 44 erfolgen zwischen 265 und 340 wesentlich schlechtere Anpassungen des linken Beins, zwischen 295 und 326 schlechtere Anpassungen des linken Armes, zwischen 333 und 348 bessere Anpassungen des linken Armes, ab 345 bessere Anpassungen des rechten Unterschenkels, ab 344 schlechtere Anpassungen der linken Schulter, im Zeitraum 356 bis 384 wesentlich schlechtere Anpassungen des linken Armes und zwischen 390 und 398 bessere Anpassungen des linken Knies.

Tabelle 3.15: Versuche 39 bis 45: Wiederholung der Versuche 31 bis 36 bei verändertem Parameter **Meßbrauschen Flächen**. Unterschiede zu den Referenzversuchen: Zwangssymmetrisierungen, Minimierungsverfahren, erweitertes Modellgradientenverfahren, **Meßbrauschen Kanten** = 12, **Kantenabstand** = **Flächenabstand** = 2, **Meßbrauschen Flächen**.

Wert ist schließlich doppelt so groß wie der andere), legen nahe, einen Parameterwert zwischen 10 und 20 auszutesten, schließlich können die genannten beiden Punkte bedeuten, daß möglicherweise zwischen den Werten 10 und 20 für **Meßbrauschen Flächen** ein Optimum für die Güte des Verfolgungsprozesses liegt. In den Versuchen 46 bis 51 werden die Versuche 31, 40, 38, 34, 43 und 45 (**Meßbrauschen Flächen** = 10) wiederholt unter Verwendung von **Meßbrauschen Flächen** = 14. Der Wert 14 wird deshalb gewählt, weil aufgrund der 15 Verschlechterungen und der 9 Verbesserungen der Versuche mit **Meßbrauschen Flächen** = 20 gegenüber den Versuchen mit **Meßbrauschen Flächen** = 10 ein eventuelles Optimum näher am Wert 10 als am Wert 20 vermutet wird. Außerdem geht der Parameter **Meßbrauschen Flächen** reziprok in die Gewichtung für den Flächenanteil der Gütefunktion Γ ein, was sogar für einen kleineren Wert als 14 spricht, nämlich $1/(\frac{1}{2}(\frac{1}{10} + \frac{1}{20})) \approx 13,33$. Für den gerundeten Wert 13 treten aber möglicherweise nur wenige Unterschiede zu den Versuchen mit dem Parameterwert 10 auf. Falls die Versuche 46 bis 51 bessere Ergebnisse bringen als die Versuche 31, 40, 38, 34, 43 und 45, so liegt zwischen den Werten 10 und 20 tatsächlich ein optimaler Wert für **Meßbrauschen Flächen**. Sollte sich eine Verschlechterung ergeben, so folgt daraus, daß zwischen den Werten kein Optimum liegt und daß für kleinere Werte als 10 wahrscheinlich bessere Anpassungsergebnisse erzielt werden. Tritt der Fall ein, daß die beiden Versuchsreihen ähnlich gute Ergebnisse liefern, so wird eine gute Verfolgung für die Werte zwischen 10 und 14 angenommen. In Tabelle 3.16 sind die Ergebnisse zusammengefaßt.

Insgesamt werden in Tabelle 3.16 bei den neuen Versuchen 3 Verbesserungen und 7 Verschlechterungen gegenüber den Versuchen 31, 40, 38, 34, 43 und 45 gezählt. Das bedeutet, daß die Wahl **Meßbrauschen Flächen** = 10 für die Fußgängerverfolgung besser

Vers.	Bildfolge	Bemerkungen
46	l_rot	Wiederholung von Versuch 31. Im Zusammenhang mit einer Verdeckung durch das rechte Bein ergeben sich zwischen 427 und 561 wesentlich schlechtere Anpassungen des linken Unterschenkels.
47	alinde	Wiederholung von Versuch 40. Im Vergleich zu Versuch 40 ergeben sich ab 195 im Zusammenhang mit einer Verdeckung durch das linke Bein wesentlich bessere Anpassungen des rechten Unterschenkels, zwischen 222 und 237 bessere Anpassungen des rechten Oberschenkels und zwischen 261 und 292 schlechtere Anpassungen von Rumpf und Becken.
48	frankm	Wiederholung von Versuch 38. Im Vergleich zu Versuch 38 erfolgen zwischen 95 und 124 im Zusammenhang mit einer Verdeckung durch das rechte Bein wesentlich schlechtere Anpassungen des linken Unterschenkels, ab 324 bessere Anpassungen der rechten Hand und dadurch ab 415 bessere Anpassungen des rechten Unterarmes, und zwischen 324 und 349 erfolgen schlechtere Anpassungen des rechten Ellenbogens.
49	g_braun	Wiederholung von Versuch 34. Zwischen 458 und 480 werden schlechtere Anpassungen des rechten Unterarmes beobachtet als in Versuch 34.
50	hof_191096	Wiederholung von Versuch 43. Im Vergleich zu Versuch 43 erfolgen zwischen 234 und 263 schlechtere Anpassungen des Rumpfes und des Beckens und zwischen 291 und 330 schlechtere Anpassungen des linken Knies.
51	s1_191096	Wiederholung von Versuch 45. Es treten keine bedeutenden Unterschiede zu Versuch 45 auf.

Tabelle 3.16: Versuche 46 bis 51: Wiederholung der Versuche 31, 40, 38, 34, 43 und 45 bei **Meßrauschen Flächen** = 14. Unterschiede zu den Referenzversuchen: Zwangssymmetrisierungen, Minimierungsverfahren, erweitertes Modellgradientenverfahren, **Meßrauschen Kanten** = 12, **Kantenabstand** = **Flächenabstand** = 2, **Meßrauschen Flächen** = 14.

geeignet ist als **Meßrauschen Flächen** = 14. Folglich liegt zwischen den Werten 10 und 20 für **Meßrauschen Flächen** kein Optimum für die Güte des Verfolgungsprozesses. Es ist anzunehmen, daß für einen Wert kleiner als 10 bessere Ergebnisse erzielt werden als bei **Meßrauschen Flächen** = 10. Daher werden in den Versuchen 52 bis 57 die Versuche 31, 40, 38, 34, 43 und 45 (**Meßrauschen Flächen** = 10) wiederholt unter Verwendung von **Meßrauschen Flächen** = 7, siehe Tabelle 3.17.

Vers.	Bildfolge	Bemerkungen
52	l_rot	Wiederholung von Versuch 31. Im Vergleich zu Versuch 31 ergeben sich zwischen 106 und 114 schlechtere Anpassungen des linken Beins, als dieses einen schwachen Kontrast zum Hintergrund aufweist.
53	alinde	Wiederholung von Versuch 40. Gegenüber Versuch 40 treten zwischen 268 und 288 bessere Anpassungen des Rumpfes sowie des rechten Oberschenkels auf und zwischen 346 und dem Bildfolgenende bei 357 erfolgen schlechtere Anpassungen des linken Oberarmes.
54	frankm	Wiederholung von Versuch 38. Im Vergleich zu Versuch 38 ergeben sich zwischen 129 und 152 im Zusammenhang mit einer Verdeckung durch das rechte Bein schlechtere Anpassungen des linken Unterschenkels, ab 321 bessere Anpassungen der rechten Hand, dadurch ab 415 bessere Anpassungen des rechten Unterarmes, und zwischen 323 und 338 erfolgen schlechtere Anpassungen des rechten Ellenbogens.
55	g_braun	Wiederholung von Versuch 34. Zwischen 328 und 357 treten bessere Anpassungen des linken Unterarmes und zwischen 417 und 474 schlechtere Anpassungen des linken Unterarmes auf als in Versuch 34.
56	hof_191096	Wiederholung von Versuch 43. Im Zeitraum 257 bis 269 ergeben sich im Vergleich zu Versuch 43 bessere Anpassungen des rechten Beins, als dieses aus der Verdeckung des linken Beines hervortritt.
57	s1_191096	Wiederholung von Versuch 45. Im Vergleich zu Versuch 45 treten zwischen 200 und 218 schlechtere Anpassungen des rechten Beins im Zusammenhang mit einer Verdeckung durch das linke Bein auf, und ab 383 erfolgen schlechtere Anpassungen des linken Knies.

Tabelle 3.17: Versuche 52 bis 57: Wiederholung der Versuche 31, 40, 38, 34, 43 und 45 bei **Meßrauschen Flächen** = 7. Unterschiede zu den Referenzversuchen: Zwangssymmetrisierungen, Minimierungsverfahren, erweitertes Modellgradientenverfahren, **Meßrauschen Kanten** = 12, **Kantenabstand** = **Flächenabstand** = 2, **Meßrauschen Flächen** = 7.

Es ergeben sich bei den Versuchen 52 bis 57 insgesamt 4 Verbesserungen und 7 Verschlechterungen gegenüber den Versuchen 31, 40, 38, 34, 43 und 45 (vgl. Tabelle 3.17). Das bedeutet, daß mit dem Wert **Meßrauschen Flächen** = 10 bessere Verfolgungser-

gebnisse erzielt werden als bei 7 oder 14. Bei **Meßrauschen Flächen** = 10 wird daher ein Optimum der Güte des Modellanpassungsprozesses angenommen. Damit soll die Optimierung der Parameter **Meßrauschen Kanten** und **Meßrauschen Flächen** abgeschlossen werden. Wie geplant, kann mit der Optimierung der Parameter **Kantenabstand** und **Flächenabstand** fortgefahren werden.

Weitere Optimierung der Parameter **Kantenabstand** und **Flächenabstand**

Nachdem sich für **Kantenabstand** = **Flächenabstand** = 2 bessere Resultate ergeben haben als bei **Kantenabstand** = **Flächenabstand** = 3 (siehe Versuche 31 bis 36), soll im folgenden untersucht werden, ob sich die Modellanpassung durch die Wahl **Kantenabstand** = **Flächenabstand** = 1 noch weiter verbessern läßt (Versuche 58 bis 63, Tabelle 3.18).

Insgesamt zählt man in Tabelle 3.18 bei den Versuchen 58 bis 63 nur 4 Verbesserungen und 17 Verschlechterungen gegenüber den Versuchen 31, 40, 38, 34, 43 und 45. Dies spricht eindeutig gegen die Verwendung von **Kantenabstand** = **Flächenabstand** = 1. Das Optimum des Verfolgungsprozesses ist daher bei **Kantenabstand** = **Flächenabstand** = 2 anzunehmen.

Angleichen des Parameters **Zylinderenden komplett**

Die einzigen Parameter, die nun noch anzugleichen sind, sind **Zylinderenden komplett** und **maximale Grauwertdifferenz für Flächenmessung**, vgl. Tabelle 3.9. Mittels der Versuche 64 bis 69 soll herausgefunden werden, ob **Zylinderenden komplett** = *ja* oder **Zylinderenden komplett** = *nein* für die Personenverfolgung besser geeignet ist. In den Versuchen wurde jeweils gegenüber den Versuchen 31, 40, 38, 34, 43 und 45 der Parameter **Zylinderenden komplett** verändert. Die Ergebnisse faßt Tabelle 3.19 zusammen.

In Tabelle 3.19 zählt man 6 Situationen, die für die Verwendung von **Zylinderenden komplett** = *ja* sprechen und 9 Situationen, die dagegen, d.h. für **Zylinderenden komplett** = *nein* verwendet werden. Vor allem zwischen den Versuchen 58 und 31 treten mehr Unterschiede auf als erwartet, schließlich steuert der Parameter **Zylinderenden komplett** nur das Aussehen derjenigen elliptischen Kegelstümpfe, deren Seitenflächen sichtbar sind (Zu jedem Modellteil wird ein Intervall angegeben, welches den zu benutzenden Bereich des definierten elliptischen Kegelstumpfes angibt. Sind die Grenzen 0 bzw. 1, so gehört die entsprechende Seitenfläche mit zum Personenmodell, andernfalls gehört sie nicht dazu): Steht der Parameter auf *ja*, so wird bei der Projektion eines Modellteils die Projektion des gesamten Umrisses einer sichtbaren Seitenfläche benutzt. Ansonsten spielt nur der Teil der Projektion eine Rolle, den man erhält, wenn man den elliptischen Kegelstumpf als Körper projiziert und nicht als Drahtmodell. Da nur wenige Modellteile sichtbare Seitenflächen besitzen und diese meist ungefähr senkrecht zur Bildebene stehen, hätte man eigentlich nur einen geringen Einfluß des Parameters

Vers.	Bildfolge	Bemerkungen
58	l_rot	Wiederholung von Versuch 31. Im Vergleich zu Versuch 31 erfolgen zwischen 130 und 156 im Zusammenhang mit einer Verdeckung des linken Beins wesentlich schlechtere Anpassungen des rechten Beins, zwischen 356 und 383 wesentlich schlechtere Anpassungen des rechten Armes, ab 365 im Zusammenhang mit einer Verdeckung durch das rechte Bein schlechtere Anpassungen des linken Unterschenkels bzw. Beins, ab 395 schlechtere Einschätzungen der Lage der rechten Schulter, zwischen 433 und 453 schlechtere Anpassungen des Rumpfes, ab 462 wesentlich schlechtere Anpassungen des Rumpfes und des Beckens, zwischen 459 und 470 schlechtere Anpassungen des rechten Knies und zwischen 461 und 477 schlechtere Anpassungen des rechten Armes.
59	alinde	Wiederholung von Versuch 40. Gegenüber Versuch 40 treten ab 195 im Zusammenhang mit einer Verdeckung durch das linke Bein wesentlich bessere Anpassungen des rechten Unterschenkels und zwischen 345 und dem Bildfolgenende bei 357 schlechtere Anpassungen des linken Oberarmes auf.
60	frankm	Wiederholung von Versuch 38. Im Vergleich zu Versuch 38 erfolgen zwischen 96 und 119 sowie zwischen 153 und 184 jeweils im Zusammenhang mit einer Verdeckung durch das rechte Bein schlechtere Anpassungen des linken Unterschenkels, zwischen 132 und 146 bessere Anpassungen des linken Unterschenkels, zwischen 325 und 354 schlechtere Anpassungen des rechten Ellenbogens, ab 325 bessere Anpassungen der rechten Hand und dadurch ab 415 bessere Anpassungen des rechten Unterarmes.
61	g_braun	Wiederholung von Versuch 34. Im Vergleich zu Versuch 34 ergeben sich zwischen 337 und 358 bessere und zwischen 410 und 474 schlechtere Anpassungen des linken Unterarmes.
62	hof_191096	Wiederholung von Versuch 43. Gegenüber Versuch 43 treten zwischen 314 und 325 schlechtere Anpassungen des rechten Beins auf, als sich dieses unter schwachem Kontrast zum Hintergrund aus der Verdeckung des linken Beins heraus bewegt, und zwischen 333 und 342 treten ebenfalls schlechtere Anpassungen des rechten Beins auf.
63	s1_191096	Wiederholung von Versuch 45. Im Vergleich zu Versuch 45 erfolgen im Halbbildzeitraum 200 bis 217 schlechtere Anpassungen des rechten Beins im Zusammenhang mit einer Ver-

Vers.	Bildfolge	Bemerkungen
		deckung durch das linke Bein, und ab Halbbild 352 erfolgen schlechtere Anpassungen des rechten Beins.

Tabelle 3.18: Versuche 58 bis 63: Wiederholung der Versuche 31, 40, 38, 34, 43 und 45 unter Verwendung von $\text{Kantenabstand} = \text{Flächenabstand} = 1$. Unterschiede zu den Referenzversuchen: Zwangssymmetrisierungen, Minimierungsverfahren, erweitertes Modellgradientenverfahren, $\text{Meßrauschen Kanten} = 12$, $\text{Meßrauschen Flächen} = 10$. $\text{Kantenabstand} = \text{Flächenabstand} = 1$.



Abbildung 3.14: Modellanpassungen von Versuch 43 (links, $\text{Zylinderenden komplett} = \text{nein}$) und Versuch 68 (rechts, $\text{Zylinderenden komplett} = \text{ja}$) zum Halbbildzeitpunkt 462 zur Demonstration der Wirkung des Parameters $\text{Zylinderenden komplett}$ auf die Projektion der linken Modellhand bei `hof_191096`. Der Parameter beeinflusst auch das Aussehen der Fußmodelle im Fersenbereich, was in der Abbildung allerdings nicht sichtbar ist. Dadurch entsteht die unterschiedliche Anpassung des rechten Fußes aufgrund der unterschiedlichen zeitlichen Entwicklung der beiden Versuche.

$\text{Zylinderenden komplett}$ auf den Verfolgungsprozeß erwartet. Abbildung 3.14 verdeutlicht die Wirkung des Parameters $\text{Zylinderenden komplett}$ auf die Projektion des Kegelstumpfmodells für die linke Modellhand bei `hof_191096`.

Angleichen des Parameters maximale Grauwertdifferenz für Flächenmessung

Nun ist nur noch der Parameter $\text{maximale Grauwertdifferenz für Flächenmessung}$ für alle Bildfolgen zu vereinheitlichen. Es soll mittels der Versuche 70 bis 75 (sie-

Vers.	Bildfolge	Bemerkungen
64	l_rot	Wiederholung von Versuch 31 unter Verwendung von Zylinderenden komplett = ja . Gegenüber Versuch 31 erfolgen jeweils im Zusammenhang mit einer Verdeckung durch das rechte Bein zwischen 366 und 384 schlechtere Anpassungen des linken Unterschenkels, zwischen 392 und 408 schlechtere Anpassungen des rechten Knies und ab 428 schlechtere Anpassungen des linken Unterschenkels. Ab 464 ergeben sich gegenüber Versuch 31 wesentlich schlechtere Anpassungen von Rumpf und Becken und zwischen 461 und 474 schlechtere Anpassungen des rechten Armes.
65	alinde	Wiederholung von Versuch 40 unter Verwendung von Zylinderenden komplett = nein . Im Vergleich zu Versuch 40 treten ab 195 im Zusammenhang mit einer Verdeckung durch das linke Bein wesentlich bessere Anpassungen des rechten Unterschenkels auf und zwischen 258 und 294 ergeben sich schlechtere Rumpf-, Kopf- und Beckenanpassungen aufgrund stärkerer Anpassungen des Rumpfes an den rechten Bildarm und den Hintergrund.
66	frankm	Wiederholung von Versuch 38 unter Verwendung von Zylinderenden komplett = nein . Gegenüber Versuch 38 erfolgen zwischen 87 und 97 schlechtere Anpassungen des linken Knies und zwischen 127 und 152 schlechtere Anpassungen des linken Unterschenkels (jeweils im Zusammenhang mit einer Verdeckung durch das rechte Bein), zwischen 324 und 350 erfolgen schlechtere Anpassungen des rechten Ellenbogens, ab 323 bessere Anpassungen der rechten Hand und dadurch ab 415 bessere Anpassungen des rechten Unterarmes.
67	g_braun	Wiederholung von Versuch 34 unter Verwendung von Zylinderenden komplett = ja . Im Zeitraum 322 bis 357 treten bessere Anpassungen des linken Unterarmes auf als in Versuch 34.
68	hof_191096	Wiederholung von Versuch 43 unter Verwendung von Zylinderenden komplett = ja . Zwischen den Halbbildern 318 und 322 werden schlechtere Anpassungen des rechten Knies beobachtet als in Versuch 43.
69	s1_191096	Wiederholung von Versuch 45 unter Verwendung von Zylinderenden komplett = ja . Im Vergleich zu Versuch 45 ergeben sich zwischen 265 und 285 bessere Anpassungen des

Vers.	Bildfolge	Bemerkungen
		rechten Beins und ab Halbbild 384 schlechtere Anpassungen des linken Knies.

Tabelle 3.19: Versuche 64 bis 69: Wiederholung der Versuche 31, 40, 38, 34, 43 und 45 bei geänderten Parameter **Zylinderenden komplett**. Unterschiede zu den Referenzversuchen: Zwangssymmetrisierungen, Minimierungsverfahren, erweitertes Modellgradientenverfahren, **Meßrauschen Kanten** = 12, **Meßrauschen Flächen** = 10, **Kantenabstand** = **Flächenabstand** = 2, **Zylinderenden komplett**.

he Tabelle 3.20) entschieden werden, welcher der Werte 25 bzw. 256 für **maximale Grauwertdifferenz für Flächenmessung** bei der Personenverfolgung besser geeignet ist, d.h. ob ein Pixelausschluß in der Flächenmessung stattfinden darf oder nicht, vgl. Seite 66. Eine weitere Optimierung bezüglich dieses Parameters ist zum gegebenen Zeitpunkt nicht geplant.

Die Auswertung der Versuche 70 bis 75 zeigt, daß bei Verwendung von **maximale Grauwertdifferenz für Flächenmessung** = 25 gegenüber **maximale Grauwertdifferenz für Flächenmessung** = 256 bei zwei der sechs Bildfolgen (**l_rot**, **frankm**) bessere Ergebnisse beobachtet werden, bei zwei Bildfolgen (**g_braun**, **s1_191096**) die Ergebnisse (im Mittel) gleich bleiben und bei den restlichen beiden Bildfolgen (**alinde**, **hof_191096**) eine starke Verschlechterung eintritt. Man zählt in Tabelle 3.20 insgesamt 6 Situationen, bei denen für den Wert 25 gegenüber 256 eine Verbesserung beobachtet wird und 12 Situationen, bei denen Verschlechterungen auftreten. Dies zeigt, daß **maximale Grauwertdifferenz für Flächenmessung** = 256 bei der modellgestützten Verfolgung verwendet werden sollte, d.h. *kein* Pixelausschluß in der Flächenmessung stattfinden sollte, vgl. Seite 66. Das Ergebnis verwundert nicht, denn falls zwei Pixel in ihrem Grauwert stark von einander abweichen, so bedeutet dies einen hohen Beitrag in der Gütefunktion Γ , d.h. hohe Kosten, welche zu minimieren sind. Solche Abweichungen treten oft bei Fehlanpassungen auf, bei welchen Kanten in den benutzten Flächenabschnitten auftreten. Mittels **maximale Grauwertdifferenz für Flächenmessung** = 256 können somit solche Fehlanpassungen erkannt werden. Ein Ausblenden der entsprechenden Pixel durch die Wahl eines kleinen Wertes für den Parameter erschwert eine robuste Verfolgung. Leider können die genannten Abweichungen auch bei Schattenbildungen oder gemusterten Kleidungsstücken auftreten, was gegen den Wert 256 spricht. Die Versuche 136 bis 141 zeigen aber, daß Probleme in diesem Zusammenhang gegenüber nicht erkannten Fehlanpassungen vernachlässigt werden sollten. Ohnehin sollten solche Effekte grundsätzlich nicht mittels einer Schwelle wie bei **maximale Grauwertdifferenz für Flächenmessung** unterdrückt werden, sondern sie sollten berücksichtigt werden, indem eine angemessenere Modellierung der Szene benutzt wird.

Zusammenfassung

Abschließend sind in Tabelle 3.21 die sich ergebenden vereinheitlichten und optimierten Parameter zusammengestellt. Die auf Seite 64 vor Durchführung der Parameterangleichung geäußerte Erwartung, daß sich die Verfolgungsergebnisse im Mittel verschlechtern, erfüllt sich nicht. Die Versuche zeigen stattdessen, daß durch die Vereinheitlichung bzw. Optimierung der Parameter *im Mittel eine Verbesserung der Verfolgungsergebnisse* erzielt werden konnte. Insgesamt ließ sich dadurch sowie durch die in Abschnitt 3.1 behandelten Maßnahmen zur Robustheitssteigerung die Anzahl an beobachteten Verfolgungsproblemen gegenüber [Wachter 97] um 35,3% reduzieren. Die aufgeführten Versuche können von nun an als Referenzversuche für weitere Untersuchungen gelten.

Vers.	Bildfolge	Bemerkungen
70	l_rot	Wiederholung von Versuch 31 bei maximale Grauwertdifferenz für Flächenmessung = 25 . Im Zeitraum 396 bis 415 beobachtet man im Zusammenhang mit einer Verdeckung durch das rechte Bein bessere Anpassungen des linken Unterschenkels als in Versuch 31.
71	alinde	Wiederholung von Versuch 65 bei maximale Grauwertdifferenz für Flächenmessung = 256 . Gegenüber Versuch 65 erfolgen zwischen 94 und 110 sowie zwischen 345 und 357 schlechtere Anpassungen des linken Oberarmes und zwischen 329 und 341 bessere Anpassungen des linken Oberarmes. In den Zeiträumen 245 bis 287 sowie 296 bis 344 und ab 347 ergeben sich gegenüber Versuch 65 wesentlich bessere Anpassungen des rechten Beins jeweils im Zusammenhang mit einer Verdeckung durch das linke Bein, zwischen 258 und 294 ergeben sich wesentlich bessere Anpassungen des Rumpf- und Beckenbereichs und zwischen 308 und 350 bessere Anpassungen des Beckens, des Rumpfes und des Kopfes.
72	frankm	Wiederholung von Versuch 66 bei maximale Grauwertdifferenz für Flächenmessung = 256 . Im Zusammenhang mit einer Verdeckung durch das rechte Bein erfolgen ab 59 gegenüber Versuch 66 wesentlich schlechtere Anpassungen des linken Beins, dadurch schlechtere Anpassungen des Rumpfes und des Beckens und zwischen 98 und 126 schlechtere Anpassungen des rechten Unterschenkels.
73	g_braun	Wiederholung von Versuch 34 bei maximale Grauwertdifferenz für Flächenmessung = 25 . Das Ergebnis ist mit dem von Versuch 34 identisch.
74	hof_191096	Wiederholung von Versuch 43 bei maximale Grauwertdifferenz für Flächenmessung = 25 . Jeweils im Zusammenhang mit einer Verdeckung durch das linke Bein ergeben sich zwischen 228 und 243 schlechtere Anpassungen des rechten Knies und des rechten Unterschenkels, zwischen 402 und 419 schlechtere Anpassungen des rechten Knies und ab 431 wesentlich schlechtere Anpassungen des rechten Beins. Zwischen 265 und 367 treten gegenüber Versuch 43 schlechtere Anpassungen des rechten Unterschenkels auf aufgrund von Freiheitsgradwertebeschränkungen des rechten Kniewinkels ab Halbbild 264, und ab 450 ergeben sich schlechtere Anpassungen des linken Unterschenkels.

Vers.	Bildfolge	Bemerkungen
75	s1_191096	Wiederholung von Versuch 45 bei maximale Grauwertdifferenz für Flächenmessung = 25. Im Zeitraum 265 bis 285 erfolgen im Vergleich zu Versuch 45 bessere Anpassungen des rechten Beins und ab Halbbild 285 schlechtere Anpassungen des linken Knies.

Tabelle 3.20: Versuche 70 bis 75: Wiederholung der Versuche 31, 65, 66, 34, 43 und 45 bei geändertem Parameter maximale Grauwertdifferenz für Flächenmessung. Unterschiede zu den Referenzversuchen: Zwangssymmetrisierungen, Minimierungsverfahren, erweitertes Modellgradientenverfahren, Meßrauschen Kanten = 12, Meßrauschen Flächen = 10, Kantenabstand = Flächenabstand = 2, Zylinderenden komplett = *nein*, maximale Grauwertdifferenz für Flächenmessung.

Parameter	Wert
Kanteninformation	EMG
Kantenabstand	2
Flächenabstand	2
Zylinderenden komplett	<i>nein</i>
Informationsnutzung	K, F
Meßrauschen Kanten	12
Meßrauschen Flächen	10
Meßrauschen Kopplung	0,1
Modellgradient, Kontrastfaktor	40
Modellgradient, sigma	3
Modellgradient, relative Mindestgröße	0,01
maximale Grauwertdifferenz für Flächenmessung	256

Tabelle 3.21: Angeglichene und optimierte Parameter bei den Bildfolgen (vgl. alte Parameter in Tabelle 3.9). EMG steht für 'erweiterter Modellgradient', K für Kanteninformation und F für Flächeninformation.

3.3 Analysierbarkeit, Zuverlässigkeit und Ausfallsicherheit

Aufgrund des komplexen Zusammenspiels zahlreicher, meist nicht überschaubarer Einzeleffekte bei der Verfolgung kommt der *Analysierbarkeit* von Ergebnissen wie auch von auftretenden Fehlern eine besondere Bedeutung zu. Das heißt, daß insbesondere geeignete grafische oder auch textuelle *Ausgaben von Zwischenergebnissen* eine wichtige Rolle bei der Analyse der Ergebnisse und der Beseitigung auftretender Fehler spielen. Die Wahl einer möglichst intuitiv zugänglichen *Form der Ein- und Ausgabedaten bzw. der Parameter* ist daher ebenso wichtig wie eine geeignete Darstellung der Zwischenergebnisse (und bei der Bildauswertung ist hier die grafische Ausgabe oft das geeignete Werkzeug). Auf der anderen Seite spielt aber auch die Analysierbarkeit von möglichen Programmfehlern bzw. die *Lokalisation von bestimmten Effekten im Programmtext* eine wichtige Rolle, schließlich muß bei vielen Verfolgungsschwierigkeiten die Ursache im Programm ausfindig gemacht werden. Ein möglichst gut strukturierter, übersichtlicher Programmcode ist daher für die Untersuchungen sehr wertvoll.

Was diese beiden Aspekte der Analysierbarkeit angeht – geeignete Ein-/Ausgabewerkzeuge sowie die Analysierbarkeit des Programmtextes – ist das zugrundeliegende Programm *Xtaro* weit fortgeschritten. Beispielsweise existiert eine Lupe zum Anzeigen der Ergebnisse sowie viele im Ergebnisfenster darstellbare Zwischenergebnisse. Zudem ist das Programm übersichtlich und sauber strukturiert programmiert. Zur Verbesserung der Ergebnisauswertung wird die Möglichkeit geschaffen, anstatt der zeilenverschränkt aufgenommenen Vollbilder im Ergebnisfenster die nach [Otte 94] tiefpaßgefilterten Halbbilder (oder zu Analyse Zwecken auch die abgeleiteten Grauwertbilder) darstellen zu können, da sie die Gegebenheiten im Anpassungsprozeß besser widerspiegeln. Dies leistete bei der Analyse wertvolle Dienste. Außerdem war ein Kontrastregler bei den Ableitungsbildern von Nöten, um die Kanten besser sichtbar zu machen.

In Sachen *Zuverlässigkeit* und *Ausfallsicherheit* zeigt sich bei den Versuchen, daß *Xtaro* gelegentlich aufgrund eines erkannten Fehlers beim Laden eines Bildes seine Arbeit unterbricht. Da die entsprechenden Bilder ordnungsgemäß vorhanden sind, kann nur ein Netzwerkfehler oder eine zu große zeitliche Netzverzögerung beim Öffnen des Bildes als Ursache für die Versuchsunterbrechung in Frage kommen. Dieser bedeutende Mangel kann durch eine sehr simple Maßnahme beseitigt werden: bei einem Lesefehler werden anstatt einem bis maximal fünf Leseversuche unternommen, bevor ein Lesefehler gemeldet wird. Dies zeigt, daß nicht immer komplexe Reaktionen auf erkannte Mängel folgen müssen.

In Bezug auf Analysierbarkeit und Ausfallsicherheit ist außerdem ein wertvolles Hilfsmittel auf Programmiererebene zu erwähnen: die Möglichkeit des verwendeten Übersetzers, Warnungen über nicht der Norm entsprechende Programmteile auszugeben. Eine nur auf den ersten Blick unbedeutende Maßnahme. Was üblicherweise von Programmierern abgeschaltet wird, entpuppt sich als wertvolle Unterstützung bei der Verbesse-

rung von Analysierbarkeit und Ausfallsicherheit. Durch Einschalten aller Warnungen des Übersetzers (neben der Bereinigung vieler hundert Warnungen ohne ernste Ursache) werden bei **Xtaro** einige Fälle möglicher undefinierter Programmzustände oder mißverständlicher Programmstellen aufgedeckt und so ein Beitrag hinsichtlich der Analysierbarkeit von Programmfehlern und der Ausfallsicherheit des Programms geleistet.

Kapitel 4

Sichtsystemgestützte Demontage

Im Einleitungskapitel wurde der Stand der Forschungen bei einem bestehenden System zur sichtsystemgestützten automatischen Demontage vorgestellt (vgl. [Tonko 97]). Im Hinblick auf das Fernziel, den Motorraum von Altfahrzeugen eines Tages vollautomatisch komponentenweise zerlegen zu können, wurde ein sichtsystembasierter Regelungsmechanismus entwickelt. Dieser konnte anhand einer exemplarischen Demontageaktion – der automatischen Entnahme der bereits von der Halterung befreiten Batterie mittels eines Greiferwerkzeuges – getestet werden. Das System ist nun im Hinblick auf das Fernziel weiter zu verbessern. Für eine solide Entwicklung müssen aufgrund der Komplexität der sich stellenden Aufgabe zum einen schrittweise die auftretenden Probleme erkannt, analysiert und mit geeigneten Mitteln beseitigt werden. Zum anderen ist die Funktionalität des Systems im Hinblick auf das Fernziel nach und nach zu erweitern. Die sich dabei zeigenden Teilprobleme wie auch die sich in natürlicher Weise stellenden Bewertungsaspekte sind dabei geeignet gegeneinander zu priorisieren, um zu jedem Zeitpunkt die richtige Wahl zu treffen und so die Entwicklung möglichst effizient vorantreiben zu können.

Zum gegenwärtigen Zeitpunkt bedeutet eine solche Analyse, daß im Hinblick auf den *Automatisierungsaspekt* vor allem eine automatische Initialisierung des bei der Regelung eingesetzten Verfolgungsprozesses benötigt wird, um eine möglichst geringe manuelle Interaktion des Programmbenutzers zu erreichen – schließlich soll in Zukunft die Demontage vollautomatisch geschehen. Da zahlreiche Schwierigkeiten bezüglich des *Robustheitsaspekts* sowie bezüglich *Zuverlässigkeit*, *Sicherheit* und *Ausfallsicherheit* des Systems beobachtet werden, auf die in [Tonko 97] nicht eingegangen werden konnte (siehe Einleitung zu Abschnitt 4.3), ist eine diesbezügliche Verbesserung unumgänglich. Natürlich dürfen auch *Erweiterungen* des Systems nicht zu kurz kommen sowie *Beschleunigungen*. Auf den Geschwindigkeitsaspekt bei der Regelung sei hier besonders hingewiesen, da er sich in Abschnitt 4.5 als sehr bedeutend auch im Hinblick auf Robustheits- und Sicherheitsaspekte erweisen wird. Die Betrachtung der Analysierbarkeit ergibt sich in natürlicher Weise während der Bearbeitung der eben genannten Aspekte. Dies zeigt, wie sich die einzelnen Bewertungsaspekte bei der Weiterentwick-

lung des Systems ergeben und welche Maßnahmen daraus folgen.

Ähnlich wie bei der Personenverfolgung sollte dem zugrundeliegenden Regelungssystem ebenfalls ein Name gegeben werden (vgl. Einleitung zu Kapitel 3). Das System soll von nun an *Visdom* heißen als Kurzform von „**V**isually **s**ervoed **d**ismantling **o**perations on (used) **m**obile vehicles“ und damit Einsatzbereich und Funktionsweise des Systems zum Ausdruck bringen. Da das System mittels künstlicher Intelligenz mit der Zeit so weit verbessert werden soll, daß es den Motorraum eines Altfahrzeuges selbständig zerlegen kann, erinnert der Name *Visdom* an das englische Wort *wisdom* (Klugheit, Weisheit).

4.1 Lokalisation zur Verbesserung der Automatisierung

Zur Automatisierung durchzuführender Aktivitäten gemäß der Definition aus Abschnitt 3.2 bildet eine automatische Initialisierung des bei der sichtsystemgestützten automatischen Demontage eingesetzten Verfolgungsprozesses ein wertvolles Hilfsmittel. Bisher stand dieser Aspekt gegenüber den Verfolgungs- und Regelungsaspekten zurück und es mußte nach jeder Änderung der Position eines Objektes oder der Kamera manuell eine Initialposition des Objektes in der Szene ermittelt werden. Die Einführung einer automatischen Lokalisation schließt die Kette zwischen Bildeinzug, Verfolgung, Regelung und dadurch realisierter Demontage. Zudem muß nicht mehr genau vorgegeben werden, wo sich ein zu demontierendes Objekt im Motorraum befindet. Es eröffnet sich in Zukunft die Möglichkeit, auch andere Fahrzeugtypen ohne genaue Positionsvorgaben zu testen oder Unfallwagen zu berücksichtigen, bei denen die betreffenden Objekte im Motorraum gegenüber ihrer werkseitigen Position verschoben sind, da die Objekte automatisch erkannt und lokalisiert werden können (zur Toleranz von Deformationen, Beschädigungen und Störeinflüssen vgl. Abschnitt 4.1.7).

Die sich stellende Hauptschwierigkeit bei der Entwicklung und Integration eines automatischen Lokalisationsverfahrens ergibt sich aufgrund der Umgebung, in welcher die Objekte eingebettet sind. Da Altfahrzeuge behandelt werden, sind Teile des Motorraumes in der Regel verschmutzt, was oft zu kontrastschwachen Bildkanten der betrachteten Objekte führt. Gleichzeitig treten im Motorraum eines Altfahrzeugs aufgrund von Kabeln, Schläuchen, anderen Motorraumteilen, Schweißnähten, Schattenwürfen oder Verschmutzungen viele Bildkanten auf, die nicht zum Objekt gehören und mit denen die Lokalisation fertig werden muß. Es stellen sich somit harte Realwelt-Anforderungen an solch ein Verfahren. Der Robustheitsaspekt spielt folglich eine herausragende Rolle. Da andererseits (im Gegensatz zur Verfolgung bei der Regelung) eine eventuell hohe Rechenzeit der Lokalisation keine Risiken für die eingesetzten Geräte in sich birgt, soll aufgrund der harten Realwelt-Anforderungen an die Robustheit ein Verfahren eingesetzt werden, das - ungeachtet der Rechenzeit - so robust wie nur möglich arbeitet. Gleichzeitig soll die Lokalisation mit so einfachen Eingabedaten wie nur möglich funktionieren, um den Interaktionsgrad niedrig zu halten. Da zur Objektverfolgung ein

CAD-Modell genügt, welches – wie die Versuche zur Verfolgung zeigen – das Objekt in der Regel gut zu beschreiben vermag, bietet es sich an, diese Information auch bei der Lokalisation zu nutzen. Da das bei der Verfolgung eingesetzte CAD-Modell hauptsächlich *Kanteninformationen* über das betrachtete Objekt liefert, wird auf dem aus dem Grauwertbild extrahierten Kantenelementbild – wie es auch bei der Verfolgung verwendet wird – operiert (vgl. Einleitung zu diesem Kapitel).

Prinzipiell gibt es zwei mögliche Verfahrensweisen zur Realisierung einer Objektlokalisierung:

- entweder man sucht im *Bildbereich* nach geeigneten Merkmalen des Objekts, ordnet diesen bestimmte Teile des Modells zu und berechnet daraus entsprechende Lagehypothesen oder
- man sucht im *Szenenbereich*, d.h. man bewertet verschiedene mögliche Modellaugen und entscheidet sich für die Lage mit der höchsten Bewertung.

Natürlich haben beide Verfahren ihre Vor- und Nachteile. Beispielsweise besteht ein Hauptproblem der ersten Verfahrensweise darin, daß die Suche signifikant von der Güte der Objektmerkmale abhängt und diese im Bildbereich sichtbar sein müssen, d.h. nicht verdeckt sein dürfen. Problematisch sind auch Bildkanten, die schlecht detektiert werden, sowie die Möglichkeit der fehlerhaften bzw. mehrdeutigen Aggregation von Kantenelementen zu Bildmerkmalen. Dies tritt besonders bei der zu behandelnden Szene zutage. Dies vermeidet die zweite Technik, jedoch ist hier als Hauptproblem die Bestimmung eines geeigneten Gütemaßes zu nennen sowie die Möglichkeit, in einem lokalen Optimum der Gütefunktion steckenzubleiben anstatt das gewünschte globale Optimum zu erreichen. Dieser Punkt entspricht bei der ersten Verfahrensweise dem Problem, Fehlzuordnungen zwischen Bild- und Modellmerkmalen vorzunehmen. Nach sorgfältiger Abwägung der Vor- und Nachteile der beiden Verfahrensweisen wird entschieden, die Szenenbereichssuche in folgender Art und Weise umzusetzen, da hierbei eine Möglichkeit gesehen wird, die meisten gravierenden Robustheitsprobleme zu lösen. Um auszuschließen, daß lediglich ein lokales Optimum der Gütefunktion gefunden wird, kommt bei der Suche bewußt ein globales Suchverfahren zum Einsatz, vgl. [Olson 98]. Insbesondere kann dadurch unterschieden werden, ob eine eventuell gefundene falsche Modellage auf eine nicht optimale Gütefunktion zurückzuführen ist oder ob ein lokales Extremum einer geeigneten Gütefunktion gefunden wurde, d.h. die Festlegung einer geeigneten Bewertungsfunktion wird deutlich unterstützt. Natürlich ist bei einer solchen Entscheidung ohne weitere Vorkehrungen mit beträchtlichen Rechenzeiten zu rechnen. Um die Dauer des Lokalisierungsverfahrens in praktikablen Grenzen zu halten, kommt daher ein *hierarchisches Abtastverfahren* im Szenenraum mit einer geeignet verallgemeinerten Gütefunktion zum Einsatz. Die Gütefunktion ermöglicht in jeder Stufe der Suche, nach dem Ort des Objekts und dessen Orientierung getrennt suchen zu können, um die Rechenzeit möglichst klein zu halten. Da eine solche Suche viele Rechenoperationen benötigt, wäre diese mit den früher verfügbaren Rechnern nur mit sehr hohem

Rechenzeitaufwand realisierbar gewesen. Die Rechenkapazität der heute verfügbaren Maschinen in Verbindung mit den genannten Optimierungen lassen jedoch einen Erfolg des Verfahrens bei akzeptabler Rechenzeit erwarten.

Bei der Wahl einer geeigneten Bewertungsfunktion ist es sinnvoll, weder Ableitbarkeit noch Stetigkeit derselben vorauszusetzen. Man ist dadurch im Gegensatz zu den bekannten lokalen Suchverfahren völlig frei in der Wahl der Bewertungsfunktion. Da nach der Lokalisationsgüte in Realweltszenen optimiert werden soll, ist dies ein besonders wichtiger Punkt. Ferner ist hervorzuheben, daß aufgrund zahlreicher Störkanten im Bild (d.h. z.T. stark ausgeprägte Kanten, die nicht zum Objekt gehören) bei oft gleichzeitig schwacher Ausprägung der Objektkanten viele lokale Minima einer Gütefunktion zu erwarten sind. Da die Gefahr lokaler Minima im vorliegenden Einsatzgebiet besonders hoch ist, bildet daher der Verzicht auf die Ausnutzung von Ableitbarkeits- bzw. Stetigkeitskriterien eine weitgehende Minimierung der vorhersehbaren Risiken. Zudem müssen bei der Suche keine zeitaufwendigen Ableitungen der Funktion berechnet und die einzelnen Modellagen können schnell bewertet werden.

Ist eine Bewertungsfunktion gefunden, für welche die genannte Suche funktioniert, so ist gleichzeitig eine Verifikationsfunktion gefunden, die dazu dienen kann, während des Verfolgungsprozesses zu überprüfen, ob das zu verfolgende Objekt noch korrekt erfaßt wird. Eine solche wird bisher in *Visdom* nicht eingesetzt, d.h. das Programm ‘bemerkt’ eine eventuelle Fehlanpassung überhaupt nicht. Diese Gelegenheit zur Robustheitssteigerung der Verfolgung ist zum Schutz der Roboter im Einsatz besonders wichtig und entsteht auf diese Weise ‘nebenbei’. Wenn der Test der im folgenden vorgestellten automatischen Objektlokalisierung erfolgreich verläuft, kann dies als Nachweis dafür angesehen werden, daß die Gütefunktion der Lokalisation als Verifikationsfunktion bei der Verfolgung geeignet ist.

Um den Suchbereich für das Objekt geschickt festlegen zu können, wird eine Zentrallage $(x_{\text{init}}, y_{\text{init}}, z_{\text{init}}, \phi_{\text{init}}, \theta_{\text{init}}, \psi_{\text{init}})^T$ des Objekts innerhalb des abzusuchenden Bereichs und für jeden der 6 Positionierungsfreiheitsgrade (x, y, z stehen stets für den Ort und ϕ, θ, ψ für die Orientierung des Objekts) eine maximale Abweichung vorgegeben ($(x_{\text{range}}, y_{\text{range}}, z_{\text{range}}, \phi_{\text{range}}, \theta_{\text{range}}, \psi_{\text{range}})^T$), um welche der entsprechende Freiheitsgrad in beide Richtungen maximal variieren darf. Dies legt den Suchbereich $[x_{\text{init}} - x_{\text{range}}, x_{\text{init}} + x_{\text{range}}] \times [y_{\text{init}} - y_{\text{range}}, y_{\text{init}} + y_{\text{range}}] \times \dots \times [\psi_{\text{init}} - \psi_{\text{range}}, \psi_{\text{init}} + \psi_{\text{range}}]$ für die Suche fest. Mit einem weiteren Parameter pro Freiheitsgrad kann ferner die initiale Anzahl an Unterteilungen für den Suchbereich angegeben werden, wodurch je Freiheitsgrad die Schrittweite der Suche, d.h. die *Granularität* $\Delta(x, y, z, \phi, \theta, \psi)^T = (\Delta x, \Delta y, \Delta z, \Delta \phi, \Delta \theta, \Delta \psi)^T$ festgelegt wird. Mit diesen Schrittweiten wird der definierte Suchbereich in einem ersten Suchschritt abgelaufen, wobei die resultierenden Modellagen jeweils bewertet werden. In weiteren Schritten wird die Suche anschließend *verfeinert*, d.h. die Granularität verkleinert.

Bei der Bewertung der einzelnen Lagen muß das Gütemaß natürlich die Granularität des Suchschrittes berücksichtigen. Dazu wird abhängig von der Granularität zu je-

der projizierten Modellkante ein Suchbereich im Bild sowie ein zur projizierten Kante kompatibler Winkelbereich für die Richtung des Grauwertgradienten ermittelt und bestimmte Kantenelemente innerhalb des Suchbereichs gezählt, die Gradientenrichtungen innerhalb des tolerierbaren Winkelbereichs repräsentieren.

4.1.1 Such- und Winkelbereiche zu den projizierten Modellkanten

Suchbereich und Winkelbereich werden zu einer projizierten Modellkante bestimmt, indem die Modellkante bei Variation der Objektlage um die halbe Granularität¹ jeweils ins Bild projiziert wird und die maximale Abweichung d_1 , d_2 der Eckpunkte der erhaltenen Projektionen sowie die maximale Abweichung φ_{\max} der Normalenrichtung der Projektionen relativ zur ursprünglichen projizierten Modellkante bestimmt werden. Genauer: Sei $(x, y, z, \phi, \theta, \psi)^T$ die zu bewertende Modellage. Die betreffende Modellkante wird unter Verwendung der Lagen

$$\left\{ \left(x + i \frac{\Delta x}{2}, y + j \frac{\Delta y}{2}, z + k \frac{\Delta z}{2}, \phi + l \frac{\Delta \phi}{2}, \theta + m \frac{\Delta \theta}{2}, \psi + n \frac{\Delta \psi}{2} \right)^T \mid i, j, k, l, m, n \in \{0, 1\} \right\}$$

projiziert und das Maximum der Abstände der so erhaltenen projizierten Modellkanten von der zu $(x, y, z, \phi, \theta, \psi)^T$ gehörenden projizierten Modellkante berechnet. Aus Geschwindigkeitsgründen wird die Menge nicht für $i, j, k, l, m, n \in \{-1, 0, 1\}$ gebildet, sondern aus Symmetriegründen davon ausgegangen, daß obige Menge etwa zu dem gleichen Ergebnis führt wie für diese i, j, k, l, m, n .² Der Winkelbereich ergibt sich anschließend aus der Normalenrichtung φ der zu bewertenden projizierten Modellkante, indem die maximale Abweichung φ_{\max} plus ein Grundrauschen von `RAUSCHEN_GRADIENTENRICHTUNG` = 6° hinzuaddiert bzw. davon abgezogen wird. Der Wert für `RAUSCHEN_GRADIENTENRICHTUNG` wurde hierbei aus den Kantenelementbildern der Objekte für geradlinige Kanten ermittelt: Da die berechnete Kantenelementrichtung bei den zu einer Kante im realen Bild gehörigen Kantenelementen bis zu $\pm 6^\circ$ (!) um einen jeweiligen Mittelwert schwankt,³ darf nicht von rein mathematischen Gegebenheiten bei den Kanten ausgegangen werden. Dieser Wert muß daher im Verfahren explizit berücksichtigt werden. Der Pixelsuchbereich wird aus der zu bewertenden projizierten Modellkante und den maximalen Abweichungen d_1 , d_2 der Kanteneckpunkte so bestimmt, daß ein Trapez entsteht. Abbildung 4.1 veranschaulicht diesen Sachverhalt.

¹Abbildung 4.2 veranschaulicht, warum hierbei die halbe Granularität sinnvoll ist.

²Bei den Forschungen zur Lokalisation ergibt sich kein Indiz darauf, daß diese Annahme nicht gerechtfertigt ist. Falls sich in Zukunft trotzdem das Gegenteil herausstellen sollte, ist natürlich $i, j, k, l, m, n \in \{-1, 0, 1\}$ anstatt $i, j, k, l, m, n \in \{0, 1\}$ in der angegebenen Menge zu verwenden.

³Das hängt mit der Rasterung der Grauwertbilder zusammen und damit, daß bei der Bildfilterung in die Gradientenrichtung abhängig von der Bildstruktur um die Bildkanten unterschiedliche Umgebungsinformation mit eingerechnet wird. Die Größe der Umgebung wird dabei durch die Größe der Filtermasken festgelegt.

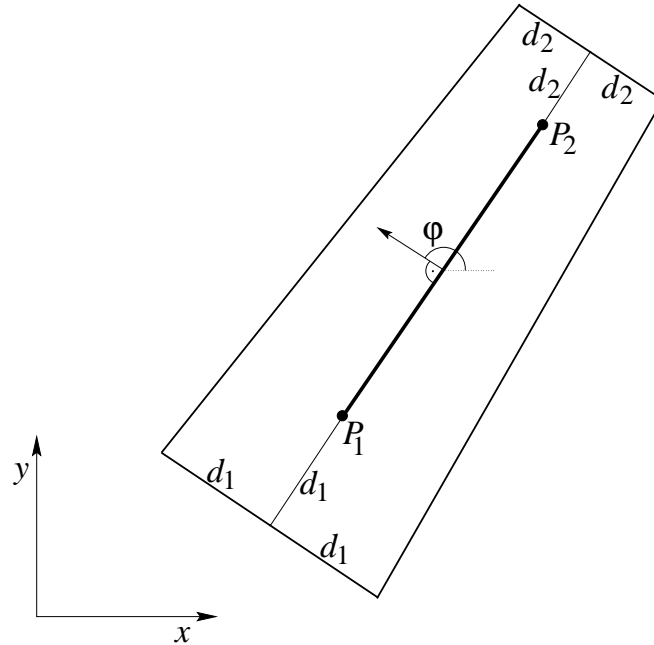


Abbildung 4.1: Suchbereich zu einer projizierten Modellkante $\overline{P_1P_2}$. φ bezeichnet die zur projizierten Kante gehörige Normalenrichtung. d_1 bzw. d_2 bezeichnen die im Text erläuterten, maximal möglichen Abweichungen der beiden Kantenendpunkte P_1 bzw. P_2 von der dargestellten Ausgangslage, abhängig von der Granularität bei der Suche.

Die Werte für d_1 und d_2 werden nach unten auf 0,5 Pixel beschränkt, damit die Güteberechnung während der Verfeinerung der Suche auch bei kleiner Granularität sinnvoll bleibt (nur dann macht die Kantenelement-Zählung innerhalb der Kantensuchbereiche Sinn).

Der so definierte Suchbereich kann zur Zählung der gewünschten, kompatiblen Kantenelemente effizient abgelaufen werden. Dazu wird zuerst der Eckpunkt des Bereichs mit minimaler y -Koordinate bestimmt und anschließend der linke Rand des Vierecks bis zur Ecke mit maximaler y -Koordinate mit dem Bresenham-Algorithmus abgelaufen, welcher – anstatt die Kanten zu zeichnen – zu jeder y -Koordinate die zugehörige (minimale) x -Koordinate speichert. Analog wird der rechte Rand des Vierecks abgelaufen und die maximalen x -Werte des Bereichs ermittelt. Anschließend kann der Suchbereich mittels dieser ermittelten Werte abgelaufen werden.

Bei der Bestimmung des Winkelbereichs und der Festlegung kompatibler Kantenelemente ist anzumerken, daß ausschließlich Winkel zwischen 0° und 180° verwendet werden. Diese Modulo-Rechnung ist erforderlich, um von der Art der Übergänge (hell nach dunkel oder dunkel nach hell) abstrahieren zu können, da das CAD-Modell keine Informationen darüber bereitstellt.

4.1.2 Bewertungsfunktion für die einzelnen Modellagen

Als Bewertungskriterium der Modellagen wird die Gütefunktion

$$f = f(x, y, z, \phi, \theta, \psi) := N \cdot \frac{P}{G}$$

verwendet. N bezeichnet dabei die Anzahl zu wertender Modellkanten, zu denen mehr als `MIN_PIXELZAHL_KANTE` := 3 Pixel im Suchbereich gefunden werden (siehe unten), und P die Gesamtsumme an Kantenelementen, die den einzelnen Kanten des projizierten Modells zugeordnet werden. G sei die Größe des Gesamtsuchbereichs, d.h. die Summe aller Suchbereichsflächeninhalte zu den projizierten Modellkanten. Der Flächeninhalt eines solchen Kantensuchbereichs kann dabei mit Hilfe der Trapezformel berechnet werden zu $(d_1 + d_2)(\overline{P_1P_2} + d_1 + d_2)$, vgl. Abbildung 4.1. Die Berücksichtigung von G verhindert, daß Modellagen systematisch bevorzugt werden, die zu einem großen Modellabbild bei der Projektion ins Bild führen. Zudem berücksichtigt G aufgrund der Berechnung der Gesamtfläche des Suchbereichs die Möglichkeit, Störkanten (Bildkanten von Bildteilen, die nicht zum Objektabbild gehören) bei der Bewertung mit einzubeziehen, d.h. bei gleichen Werten für N und P wird eine Modellage um so höher bewertet, je kleiner (G und damit) die Möglichkeit von Fehlzuordnungen ist. Es sei bemerkt, daß G gegen die Gesamtlänge der projizierten Modellkanten konvergiert, wenn die Granularität gegen Null geht.⁴ Die Einrechnung von N in die Güte vermeidet, daß das Modell bevorzugt an Bildteile angepaßt wird, wo besonders viele Kantenelemente detektiert wurden, wie zum Beispiel an Bildteile mit mehreren parallelen Bildkanten, die möglicherweise nicht zu dem zu lokalisierenden Objekt gehören. Anders ausgedrückt trägt N dazu bei, Bildstrukturen zu suchen, die dem Abbild des projizierten Objekts möglichst ähnlich sind, und dazu gehört, daß möglichst vielen Modellkanten möglichst viele Kantenelemente des Bildes zugeordnet werden.

Wie Versuche zeigen, hat N erheblichen Einfluß auf die berechnete Güte, weshalb nur hinreichend wichtige Objektkanten gezählt werden dürfen. Daher werden bei N nur die Modellkanten gezählt, zu denen mindestens `MIN_PIXELZAHL_KANTE` = 3 Pixel gefunden werden (unterhalb dieses Wertes kann man schwer davon sprechen, eine Gerade gefunden zu haben, welcher maximal viele Kantenelemente des Suchbereichs zugeordnet wurden). Außerdem werden bei N und P nur die projizierten Modellkanten gewertet, die länger sind als $(d_1 + d_2)/8$, vgl. Abb. 4.1. Dies bedeutet in natürlicher Weise bei größeren Suchschritten (d.h. bei den anfänglichen Suchschritten mit grober Granularität) eine Ausblendung zu feiner Details, deren Berücksichtigung dort zum Auffinden der Groblage des Objekts keinen Sinn macht, und eine Hinzunahme immer feinerer Details von Suchschritt zu Suchschritt.

Bei der Zählung der Kantenelemente P innerhalb der Suchbereiche wird die potentiell höchste Bewertung der einzelnen projizierten Modellkanten bestimmt, die in folgen-

⁴Versuche, bei denen im Gütemaß die Gesamtlänge an projizierten Modellkanten anstatt G verwendet wird, liefern schlechtere Modellagenbewertungen. Insbesondere ist in diesem Fall die Güte um so größer, je größer das projizierte Modell ist.

den Suchstufen maximal auftreten kann. Ist die ermittelte Pixelzahl innerhalb eines Suchbereichs größer als die Länge der projizierten Modellkante (in Pixeln gemessen), so wird daher die Pixelzahl auf diese Länge begrenzt (schließlich macht es keinen Sinn, einer projizierten Modellkante mehr Pixel zuzuordnen, als sie selbst lang ist). Andernfalls würde man die Modellkante in der Güte übergewichten und dadurch niedrige Pixelzahlen von projizierten Modellkanten mit wenigen zugeordneten Kantenelementen kompensieren. Die beim Ablaufen aller Suchbereiche erhaltene Summe P an Pixeln dient als wichtigster Faktor beim Gütemaß.

Den Versuchen ist zu entnehmen, daß sich hohe Werte von f desto besser mit visuell als gut beurteilten Modellanpassungen decken, je kleiner die Suchbereiche zu den einzelnen projizierten Modellkanten sind (klar wegen der Kanten im Bild, die nicht zum Objekt gehören, siehe z.B. in Abbildung 1.3 die Bildkanten in der rechten oberen Ecke, die nahezu parallel zu der am nächsten liegenden Batteriekante sind, oder die Kanten links von der Batterie).

4.1.3 Verfeinerung der Suche

Bei der Verfeinerung der Suche, d.h. bei sukzessiver Verkleinerung der Granularität in den einzelnen Suchschritten, darf natürlich nicht davon ausgegangen werden, daß zu der am höchsten bewerteten Modellage auch die tatsächlich beste Modellanpassung gehört. Schließlich muß das Suchverfahren stets die Möglichkeit berücksichtigen, daß manchen Modellkanten insbesondere bei grober Granularität Bildstrukturen zugeordnet werden können, die nicht zum Objektabbild gehören. Daher wird davon ausgegangen, daß sich *eine der* am besten bewerteten Modellagen mit einer bzgl. der gewählten Granularität guten Anpassung deckt (pessimistischer Algorithmus). Zur Vermeidung von Fehllokalisationen wird die Entscheidung, welche der Modellagen zum gewünschten Ergebnis führt, in den jeweils folgenden Suchschritt transferiert. Zur Begrenzung der Größe des Suchbereichs auf der einen Seite und zur Begrenzung der Gesamtrechenzeit auf der anderen Seite wird in der ersten Iteration eine *Grobsuche* mit hinreichend kleiner Granularität realisiert und in weiteren Iterationen eine *Feinsuche* mit möglichst wenigen zu bewertenden Modellagen, d.h. mit jeweils maximaler Granularität (die natürlich abhängt von der Granularität der jeweils vorhergehenden Iteration). Bei der Grobsuche hat es sich herausgestellt, daß eine Suche im 6-dimensionalen Raum erforderlich ist, um die Größe der Kantensuchbereiche im Bild zu begrenzen. Aus Rechenzeitgründen besteht jeder Iterationsschritt der Feinsuche dagegen aus einer 3-dimensionalen Suche nach der Orientierung des Objekts sowie einer anschließenden 3-dimensionalen Suche nach dem Ort.

Wird die initiale Granularität für die drei Orientierungsfreiheitsgrade bei der Orientierungssuche klein gewählt, werden nur Kantenelemente gezählt, die hinsichtlich der repräsentierten Gradientenrichtung relativ genau zu den projizierten Modellkanten passen. Die dabei vorzusehende Ortsunschärfe kann dann bei der anschließenden Suche

nach dem Ort ausgeglichen werden. Bei den als zeitaufwendig bekannten trigonometrischen Funktionen zur Sinus-, Cosinus- und Arcustangensberechnung sowie bei der Wurzelberechnung werden effizient auswertbare Funktionsdarstellungen verwendet, die mit Tabellenzugriffen arbeiten.

Bei der Verfeinerung der Suche, d.h. beim Übergang von einer Iteration zur nächsten können für die am besten bewerteten Modellagen die in der folgenden Iteration weiter zu prüfenden Modellagen, die Granularität sowie die bei der Bestimmung der Such- und Winkelbereiche zu den einzelnen projizierten Modellkanten relevanten Größen (halbe Granularität, vgl. S. 96) gemäß Abbildung 4.2 einfach berechnet werden.

Die sukzessive Halbierung des abzusuchenden Bereichs für die einzelnen Freiheitsgrade führt – im Gegensatz zur Grobsuche, bei der je Freiheitsgrad mehr als eine Unterteilung erfolgt – bei der Feinsuche zu möglichst wenigen Abtastungen des Suchbereichs

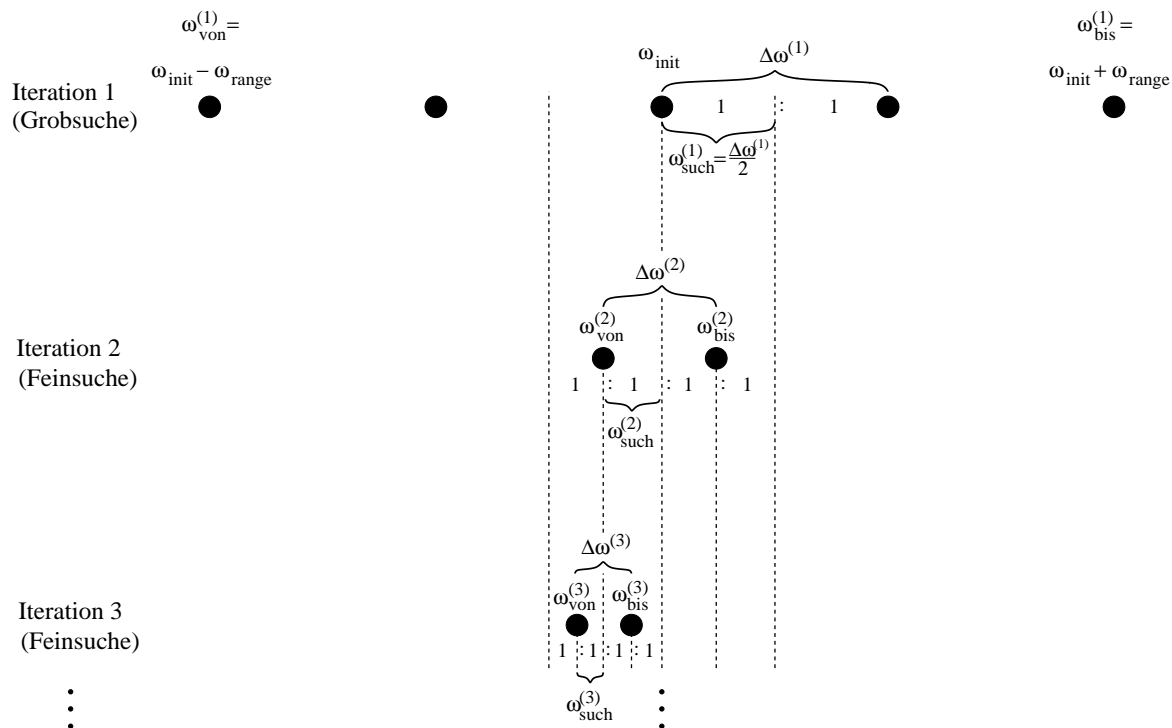


Abbildung 4.2: Veranschaulichung der bei den Iterationen i der hierarchischen Suche zu bewertenden Modellagen \bullet , den verwendeten Granularitäten $\Delta\omega^{(i)}$ und die bei der Bestimmung der Such- und Winkelbereiche zu den einzelnen projizierten Modellkanten relevanten Größen $\omega_{such}^{(i)} = \Delta\omega^{(i)}/2$ (vgl. S. 96) sowie der Zusammenhang zwischen diesen Werten ($1 : 1$ kennzeichnet jeweils Unterteilungen im Mittelpunkt). ω steht für den entsprechenden Freiheitsgrad x, y, z, ϕ, θ bzw. ψ der gesuchten Modellposition. $\omega_{von}^{(i)}$ und $\omega_{bis}^{(i)}$ bezeichnen jeweils den Anfang und das Ende der Suche bezüglich ω . Die Abbildung zeigt jeweils exemplarisch das weitere Vorgehen in der Folgeiteration $i \geq 2$ bezüglich einer weiter zu untersuchenden Modellage der vorhergehenden Iteration.

innerhalb eines Suchschrittes. Die Suche endet, wenn die Positionierungsgenauigkeit des projizierten Modells im Bild unter $\text{PIXELGENAUIGKEIT} := 1 \text{ Pixel}$ fällt.⁵ Innerhalb einer Iteration erfolgt zuerst die Winkelsuche für die Winkelfreiheitsgrade entsprechend der in Abbildung 4.2 veranschaulichten Zusammenhänge, wobei bei den fixen Ortsfreiheitsgraden ω mit maximalem $\omega_{\text{such}}^{(i)}$ gearbeitet wird, d.h. $\omega_{\text{such}}^{(i)} = \omega_{\text{range}}$ für $i = 1$ bzw. $\omega_{\text{such}}^{(i)} = \omega_{\text{such}}^{(i-1)}$ für $i > 1$. Anschließend erfolgt die 3D-Ortssuche für die Ortsfreiheitsgrade gemäß der Zusammenhänge von Abbildung 4.2, wobei nun bei den fixen Winkelfreiheitsgraden ω – sichergestellt durch die vorausgehende Winkelsuche – mit dem gleichen $\omega_{\text{such}}^{(i)}$ gearbeitet werden kann wie bei der vorausgehenden Winkelsuche (vgl. Abbildung 4.2).

Nach jeder Teilsuche (Orientierungs- oder Ortssuche) werden die $p\%$ am besten bewerteten Modellagen in der folgenden Teilsuche weiter untersucht und die übrigen verworfen. Es bietet sich aus Effizienzgründen an, p nicht konstant anzusetzen, sondern von Teilsuche zu Teilsuche kleiner werden zu lassen. In der letzten Teilsuche der letzten Iteration sollte p letztendlich so klein sein, daß nur noch eine Modellage ausgewählt wird, schließlich spielt nur noch die Modellage mit der höchsten Bewertung eine Rolle. Mit der Konvention, daß auch für p nahe 0 mindestens eine Modellage ausgewählt wird, soll dieser Sachverhalt durch $p \approx 0\%$ repräsentiert werden. Es stellt sich die Frage, nach welchen Kriterien p zu wählen ist, um möglichst wenige Modellagen bewerten zu müssen, ohne wichtige Modellagen zu vernachlässigen. Bei der Beantwortung dieser Frage spielt es eine wichtige Rolle, wie stark die Möglichkeit von Fehlzuordnungen von Teilsuche zu Teilsuche abnimmt. Die Möglichkeit von Fehlzuordnungen hängt ab von der Größe von G , welche wiederum von den einzelnen d_1, d_2 der projizierten Modellkanten abhängt. Die projizierte Modellkante mit dem größten Wert für d_1 bzw. d_2 hat dabei den größten Einfluß auf die Fehlzuordnungswahrscheinlichkeit. Sei daher d_{max} das Maximum der d_1, d_2 aller projizierten Modellkanten zu der aktuell untersuchten Modellage. d_{max} ist ein Maß für die Fehlzuordnungswahrscheinlichkeit und dient daher als Grundlage für die Bestimmung von p , d.h. p wird in Abhängigkeit von dieser Größe gewählt. Wenn d_{max} unter $\text{PIXELGENAUIGKEIT} = 1 \text{ Pixel}$ fällt, sollte $p \approx 0\%$ sein.

$\overline{P_1 P_2}$ sei eine projizierte Modellkante bezüglich der aktuellen Modellage mit $d_1 = d_2 = d_{\text{max}}$. Der zugehörige Kantensuchbereich hat den zur Fehlzuordnungswahrscheinlichkeit proportionalen Flächeninhalt

$$A_1 := 2d_{\text{max}} \cdot (\overline{P_1 P_2} + 2d_{\text{max}}) .$$

Bei Verdopplung von d_{max} vergrößert sich der Kantensuchbereich auf die Fläche

$$A_2 := 4d_{\text{max}} \cdot (\overline{P_1 P_2} + 4d_{\text{max}}) ,$$

d.h. bei Verdopplung von d_{max} erhöht sich die Suchfläche und damit p auf das s -fache

⁵Ein kleinerer Wert für PIXELGENAUIGKEIT macht keinen Sinn, da die gezählten Kantenelemente von den zugehörigen Geraden einen Abstand von $\text{abst}_{\text{max}} = 1,0 \text{ Pixel}$ haben dürfen, vgl. S. 103.

mit

$$\begin{aligned} s &:= \frac{A_2}{A_1} = 2 \frac{(\overline{P_1 P_2} + 2d_{\max}) + 2d_{\max}}{\overline{P_1 P_2} + 2d_{\max}} = 2 \left(1 + \frac{2d_{\max}}{\overline{P_1 P_2} + 2d_{\max}} \right) \\ &= 2 \left(1 + \frac{1}{\frac{\overline{P_1 P_2}}{2d_{\max}} + 1} \right) \in (2, 4] . \end{aligned}$$

p kann daher angesetzt werden zu

$$p = a \cdot d_{\max} + b \cdot d_{\max}^2 ,$$

wobei a und b noch geeignet zu bestimmen sind. Je näher s bei 2 liegt, desto mehr Modellagen werden in einer Folgeiteration der Suche weiter berücksichtigt. Der Suchalgorithmus ist daher ‘am vorsichtigsten’, wenn $b = 0$ gewählt wird. Aus diesem Grund wird mit $b = 0$ gearbeitet und a so gewählt, daß mit einem initialen Wert p_{init} für p die Menge weiter zu untersuchender Modellagen für die Winkelsuche der ersten Iteration vorgegeben werden kann. Auf diese Weise kann auf die anzunehmende, grundlegende Fehlzuordnungswahrscheinlichkeit Einfluß genommen werden. Auf der anderen Seite kann der Parameter p_{init} interpretiert werden als Vermittler zwischen globaler und lokaler Optimierung der Gütefunktion: Für einen großen Wert ergibt sich die ursprünglich vorgesehene globale Optimierung der Gütefunktion. Da die Gütefunktion mit der Anpassungsgüte korreliert ist, schlagen zu große Werte unnötig auf die Rechenzeit - ohne Robustheitsgewinn. Je kleiner p_{init} gewählt wird, desto ‘lokaler’ funktioniert die Optimierung der Gütefunktion und desto kleiner ist die Rechenzeit. p_{init} dient also der Verringerung der Rechenzeit. Zu kleine Werte bedeuten ein Risiko, die Suche zu stark lokal einzugrenzen. Im Zweifelsfall ist p_{init} lieber zu groß zu wählen.

Ein Wechselpuffer dient dazu, alle Modellagen samt Bewertung der letzten Teilsuche zu sammeln, die Ermittlung der $p\%$ besten Modellagen zu ermöglichen und schließlich in einem zweiten Speicherbereich die Ergebnisse der folgenden Teilsuche aufzunehmen. Die Bedeutung des ersten und zweiten Pufferspeicherbereichs wechselt dabei natürlich bei den Teilsuchen ab.

4.1.4 Zählung der geeigneten Kantenelemente in den Suchbereichen

Bei der Bewertung der projizierten Modellkanten, d.h. der Zählung der Kantenelemente innerhalb der Kantensuchbereiche werden nur die kompatiblen Kantenelemente gezählt, die nahezu auf einer einzigen Geraden liegen (genau auf der Geraden werden i.A. aufgrund der Bildraasterung nur die wenigsten liegen). Dabei ist diejenige Gerade auszuwählen, der maximal viele kompatible Kantenelemente zugeordnet werden können. Das Verfahren ist dabei so eingestellt, daß Kantenelementzüge zu einzelnen

Bildkanten möglichst komplett erfaßt und gleichzeitig möglichst wenige Störpixel mit eingerechnet werden. Zur effizienten Durchführung wird die im folgenden beschriebene Variante der Hough-Transformation eingesetzt. Daß es sinnvoll ist, die Suche im Szenenraum mit einer Hough-Transformation zu kombinieren, wird in [Olson 99] belegt (auch wenn in der angegebenen Literaturstelle ein anderer Weg der Kombination beschrrieben wird).

Eine Gerade kann eindeutig beschrieben werden durch die Richtung α der Normalen und den kürzesten Abstand d zu einem festen Punkt (= Ursprung des Koordinatensystems). Man denke sich den nach Kantenelementen abzusuchenden Kantensuchbereich überdeckt von Geraden beliebiger Parameter (α, d) (geeignet diskretisiert). Zu jedem bezüglich der Richtung der projizierten Modellkante kompatiblen Kantenelement innerhalb des Kantensuchbereichs existiert eine bestimmte Teilmenge dieser Geraden, die von dem Kantenelement einen Maximalabstand von $abst_{\max} := 1$ Pixel haben und deren Richtung kompatibel zu der Gradientenrichtung des Kantenelements ist. Zu all diesen Geraden aus der Teilmenge wird abhängig von (α, d) der jeweils zugehörige Eintrag eines zweidimensionalen Zählfeldes erhöht. Nach Abarbeitung aller kompatiblen Kantenelemente des Suchbereichs enthält schließlich jeder Eintrag des Zählfeldes die Anzahl an Kantenelemente, welche von der zugehörigen Geraden einen Höchstabstand von $abst_{\max}$ haben. Die Gerade mit maximalem Zählfeldeintrag ist schließlich die Gerade mit maximaler Anzahl an zugeordneten Pixeln. Der Feldeintrag gibt die gesuchte Pixelzahl, d.h. die (noch nicht normierte) Bewertung der projizierten Modellkante an. Der Parameter $abst_{\max}$ ist notwendig, um zu gewährleisten, daß alle Kantenelemente der Bildkanten vollständig erfaßt werden, schließlich sind die Kantenelemente von Bildkanten im Allgemeinen *nur ungefähr* kollinear angeordnet.

Sei \mathbf{p} die Position (Ortsvektor) eines (bzgl. der projizierten Modellkante kompatiblen) Kantenelements mit Gradientenrichtung α und $e \in [-abst_{\max}, abst_{\max}]$. Zur Bestimmung aller relevanten Geraden (β, d) bezüglich \mathbf{p} kann β den Bereich $[\alpha - 6^\circ, \alpha + 6^\circ]$ (diskret) ablaufen. Zu jeder Geraden mit Normalenrichtung β sind die Geraden mit Abstand e bestimmt durch die Gleichung

$$\mathbf{n}^T \mathbf{p} - d = e, \quad \text{wobei} \quad \mathbf{n} = \begin{pmatrix} \cos \beta \\ \sin \beta \end{pmatrix}.$$

Es muß daher lediglich

$$d = \mathbf{n}^T \mathbf{p} - e$$

für alle e aus dem angegebenen Bereich berechnet werden, um die gewünschten Geraden (β, d) zu erhalten. Die Bewertung projizierter Modellkanten kann also wie folgt implementiert werden:

Algorithmus Bewertung einer projizierten Modellkante

Eingabe:

Suchbereich einer projizierten Modellkante

```

Initialisiere feld[*,*] := 0;
für alle zur Kante kompatiblen Kantenelemente  $(\mathbf{p}, \alpha)$  des Suchbereichs
  für alle  $\beta \in \{\alpha - 6^\circ, \alpha - 5^\circ, \dots, \alpha + 6^\circ\}$ 
     $\mathbf{n} := \begin{pmatrix} \cos \beta \\ \sin \beta \end{pmatrix}$ ; /* Normalenvektor */
    für alle  $d \in \{\mathbf{n}^T \mathbf{p} - abst_{\max}, \mathbf{n}^T \mathbf{p} - abst_{\max} + 1, \dots, \mathbf{n}^T \mathbf{p} + abst_{\max}\}$ 
      feld[ $d, \beta$ ] := feld[ $d, \beta$ ] + 1;
  Suche das  $(d, \beta)$  mit maximalem feld[ $d, \beta$ ];

```

Ausgabe: Kantenbewertung feld[d, β].

Zur Beschleunigung der Suche wird diese Bewertung der projizierten Modellkanten optimiert. Dazu wird ein Großteil der in obigem Algorithmus dargestellten Funktionalität aus der eigentlichen Suche herausgezogen und in einer Vorverarbeitung der Suche berechnet. Die Vorverarbeitung erledigt dabei die Berechnung der relevanten Geraden zu den einzelnen Kantenelementen des Gesamtbildes und liefert die Zeiger auf die entsprechenden, zugehörigen Feldeinträge feld[*,*]. Bei der eigentlichen Suche müssen anschließend nur noch die entsprechenden Feldeinträge, adressiert durch die vorberechneten Zeiger, erhöht werden.

4.1.5 Multiokulare Lokalisation

Die entwickelte Objektlokalisierung kann in intuitiver und wirkungsvoller Weise einfach auf den multiokularen Fall erweitert werden.

Seien k kalibrierte Kameras gegeben, die jeweils ein Bild des zu lokalisierenden Objektes liefern. Zu jeder Kamera kann wie bisher N , P , G und d_{\max} berechnet werden. Aus den so erhaltenen Werten N_i , P_i , G_i und $d_{\max,i}$ ($i = 1, \dots, k$) können N_Σ , P_Σ , G_Σ bzw. $d_{\max,\Sigma}$ für den multiokularen Fall sowie die multiokulare Lagebewertung f wie folgt berechnet werden:

$$f = N_\Sigma \cdot \frac{P_\Sigma}{G_\Sigma} \quad , \quad (4.1)$$

wobei

$$N_\Sigma = \sum_{i=1}^k N_i \quad , \quad P_\Sigma = \sum_{i=1}^k P_i \quad , \quad G_\Sigma = \sum_{i=1}^k G_i \quad . \quad (4.2)$$

Ferner

$$d_{\max,\Sigma} = \max_{i=1,\dots,k} \{d_{\max,i}\} \quad . \quad (4.3)$$

Die multiokulare Lagebewertung f wird also *nicht* direkt aus den monokularen Einzelbewertungen $f_i = N_i P_i / G_i$ ($i = 1, \dots, k$) der Kameras (etwa mit Hilfe des arithmetischen Mittels) berechnet, um zu vermeiden, daß monokulare Einzelbewertungen

mit wenigen korrekt zugeordneten Bildkanten mit gleicher Gewichtung ins Gütemaß eingehen wie welche mit vielen (z.B. bei Modellagen, die zu einer guten Anpassung im einen und einer schlechten Anpassung in einem anderen Kamerabild führen).

4.1.6 Berücksichtigung von a-priori Wissen zur Objektlage

Um eine hohe Effizienz des vorgestellten Lokalisationsverfahrens zu erzielen sowie um den Suchraum und damit die Suchzeit auf ein notwendiges Maß zu begrenzen, wird die Abtastung im Szenenraum bezüglich eines *objektzentrierten Objektkoordinatensystems* durchgeführt, d.h. bzgl. eines lokalen Objektkoordinatensystems, dessen Ursprung mit dem Schwerpunkt des Objektmodells zusammenfällt. Dies sichert eine weitgehende Entkopplung zwischen Objektposition und -orientierung.

Die Verwendung dieses Koordinatensystems ermöglicht gleichzeitig in einfacher Weise a-priori Wissen zur Objektlage zu berücksichtigen und so den Suchraum auf das benötigte Maß zu begrenzen. Beispielsweise befindet sich das Schiffsteil (siehe Abschnitt 4.1.7) auf einem rollbaren Wagen stets mit der Grundfläche parallel zur Bodenebene. Folglich sind die Rotation gegen die Bodenebene sowie die Höhe über dem Boden stets fest. Durch Wahl eines geeigneten \mathbf{x}_{init} zur Definition der festen Lageparameter (z.B. durch Lokalisation des Objektes bzgl. aller 6 Freiheitsgrade in einer Position inmitten des Suchbereichs) kann bei der Objektlokalisierung die Suche auf weniger als 6 Freiheitsgrade reduziert werden, indem die entsprechenden Parameter zur Ausdehnung der Suchbereichsgröße auf Null gesetzt werden (z.B. beim Schiffsteil $z_{\text{range}} = \theta_{\text{range}} = \psi_{\text{range}} = 0$, die eigentliche Suche umfaßt nur noch 3 Freiheitsgrade).

4.1.7 Ergebnisse

Da explizit die Möglichkeit von Fehlanpassungen im Verfahren berücksichtigt, Artefakte realer Kanten im Bild einbezogen und Orts- wie auch Richtungsinformation der Kantenelemente verwendet worden sind sowie lokale Minima der Gütefunktion weitgehend ausgeschlossen worden sind (vgl. Seite 102), kann von der Lokalisation erwartet werden, daß sie bei den meisten Objekten trotz komplexem Hintergrund robust arbeitet. Erste Tests zeigen, daß das Gütemaß gut zur Bewertung der Lagen geeignet zu sein scheint.

Durch die sukzessive Verringerung der Granularität und damit auch der Kantensuchbereiche erfolgt im Verlauf der Lokalisation eine 'selbst zuziehende' Suche. Fehler zwischen Modellprojektion und Bild werden folglich mindestens so lange toleriert, wie die damit verbundene Abweichung im Bild unterhalb der Größe der Kantensuchbereiche liegt. Solche Fehler können durch Ungenauigkeiten im Modell, deformierte oder beschädigte Objekte des Motorraumes von Unfallwagen, Bildverzerrungen oder fehlerhafte bzw. un-

genaue Kamerakalibrierungen zustande kommen und sind oft nicht zu vermeiden. Als Konsequenz kann von dem vorgestellten Lokalisationsverfahren erwartet werden, daß es gegenüber solchen Fehlerquellen so weit tolerant arbeitet, daß Anpassungen geliefert werden, die mindestens genau sind bezüglich der mit solchen Fehlern einhergehenden Ungenauigkeit im Bild.

Die Anwendung der Lokalisation auf sämtliche verwendeten Objekte führt in allen Fällen zu genauen Anpassungen (genau bzgl. weniger Pixel). Die Erwartungen an die Robustheit werden damit erfüllt. In den Abbildungen 4.3 bis 4.5 sind zur Illustration die Lokalisationsergebnisse aller betrachteten Objekte für die relevanten mono- und binokularen Fälle dargestellt. Die monokulare Lokalisation des Schiffsteils zeigt Abbildung 4.24 auf Seite 141 (oben). Zum unteren in Abbildung 4.3 dargestellten Objekt sei auf die Beschreibung in Abschnitt 4.4.2 verwiesen. Zu dem kugelförmigen Unterdruckbehälter und dem Schiffsteil siehe Abschnitt 4.4.3.

Um die Objektlokalisierung neuen Herausforderungen gegenüberzustellen und so einen noch besseren Eindruck von dem tatsächlich erreichten Grad der Robustheit erlangen zu können, wird das Verfahren angewandt auf 6 Schiffsteilbilder der Größe 1296×1030 Pixel im *Vollbildformat*, die von einer Schiffswerft mit einer hochauflösenden Kamera aufgenommen und freundlicherweise zur Verfügung gestellt worden sind (zum Vergleich: oben referierte Bilder im *Halbbildformat* haben eine Auflösung von 584×264 Pixel; da die Verfolgung mit Halbbildern arbeitet, wurde auch bei der Lokalisation mit solchen gearbeitet). Es ist damit zu rechnen, daß hierbei Lokalisationsungenauigkeiten auftreten können. Wie die erzielten Ergebnisse jedoch zeigen, erfolgt die Anpassung bezüglich einer Genauigkeit von maximal 10 Pixeln, was aufgrund der Auflösung im Rahmen bleibt, siehe Abbildungen 4.6 bis 4.8.

Tabelle 4.1 faßt die bei den dargestellten Lokalisationen verwendeten Parameter zusammen. p_{init} ist bei der monokularen Batterielokalisation deshalb so hoch, um ein lokales Minimum der Gütefunktion aufgrund einer ungünstigen Konfiguration der Kantenelemente in der Umgebung der Batterie zu vermeiden, vgl. Seite 102. Die Lokalisationszeiten bewegen sich auf einer SUN Sparc Ultra II mit zwei 300 MHz-Prozessoren zwischen knapp 3 Minuten beim Doppelklotz (siehe zu diesem Objekt Abschnitt 4.4.2), 20 Minuten beim Schiffsteil im Bild der hochauflösenden Kamera bzw. über 45 Minuten bei der ungünstigen Konfiguration in Form der monokularen Batterielokalisation.

Da die Objektlokalisierung robust funktioniert, ist die zugrundeliegende Gütefunktion als tauglich für die Bewertung guter Anpassungen nachgewiesen. Sie sollte damit geeignet sein als Verifikationsfunktion während der Verfolgung (vgl. Kapitel 6).

Lokalisation	p_{init}	$\mathbf{x}_{\text{range}}, \Delta \mathbf{x}$
Batterie	17%	$(50\text{mm}, 50\text{mm}, 20\text{mm}, 45^\circ, 15^\circ, 15^\circ)^T$, $(25\text{mm}, 25\text{mm}, 20\text{mm}, 15^\circ, 15^\circ, 15^\circ)^T$,
Luftfilter	7%	$(30\text{mm}, 30\text{mm}, 30\text{mm}, 5^\circ, 5^\circ, 5^\circ)^T$, $(15\text{mm}, 15\text{mm}, 15\text{mm}, 5^\circ, 5^\circ, 5^\circ)^T$,
kugelförmiger Behälter	7%	$(30\text{mm}, 30\text{mm}, 30\text{mm}, 5^\circ, 5^\circ, 5^\circ)^T$, $(15\text{mm}, 15\text{mm}, 15\text{mm}, 5^\circ, 5^\circ, 5^\circ)^T$,
Doppelklotz	10%	$(80\text{mm}, 80\text{mm}, 0\text{mm}, 180^\circ, 0^\circ, 0^\circ)^T$, $(8,\overline{8}\text{mm}, 8,\overline{8}\text{mm}, 0\text{mm}, 8,\overline{18}^\circ, 0^\circ, 0^\circ)^T$,
Schiffsteil, Standardkamera	7%	$(300\text{mm}, 300\text{mm}, 0\text{mm}, 180^\circ, 0^\circ, 0^\circ)^T$, $(75\text{mm}, 75\text{mm}, 0\text{mm}, 10^\circ, 0^\circ, 0^\circ)^T$,
Schiffsteil, hochauflösende Kamera	5%	$(400\text{mm}, 400\text{mm}, 0\text{mm}, 180^\circ, 0^\circ, 0^\circ)^T$, $(100\text{mm}, 100\text{mm}, 0\text{mm}, 10^\circ, 0^\circ, 0^\circ)^T$,
Batterie	7%	$(50\text{mm}, 50\text{mm}, 20\text{mm}, 45^\circ, 15^\circ, 15^\circ)^T$, $(25\text{mm}, 25\text{mm}, 20\text{mm}, 15^\circ, 15^\circ, 15^\circ)^T$,
Luftfilter	7%	$(30\text{mm}, 30\text{mm}, 30\text{mm}, 5^\circ, 5^\circ, 5^\circ)^T$, $(15\text{mm}, 15\text{mm}, 15\text{mm}, 5^\circ, 5^\circ, 5^\circ)^T$,
kugelförmiger Behälter	7%	$(30\text{mm}, 30\text{mm}, 30\text{mm}, 5^\circ, 5^\circ, 5^\circ)^T$, $(15\text{mm}, 15\text{mm}, 15\text{mm}, 5^\circ, 5^\circ, 5^\circ)^T$,
Schiffsteil, Standardkameras	7%	$(300\text{mm}, 300\text{mm}, 0\text{mm}, 180^\circ, 0^\circ, 0^\circ)^T$, $(75\text{mm}, 75\text{mm}, 0\text{mm}, 10^\circ, 0^\circ, 0^\circ)^T$,

Tabelle 4.1: Lokalisationsparameter aller monokularen (oberer Teil der Tabelle) und binokularen Lokalisationen (unterer Teil).

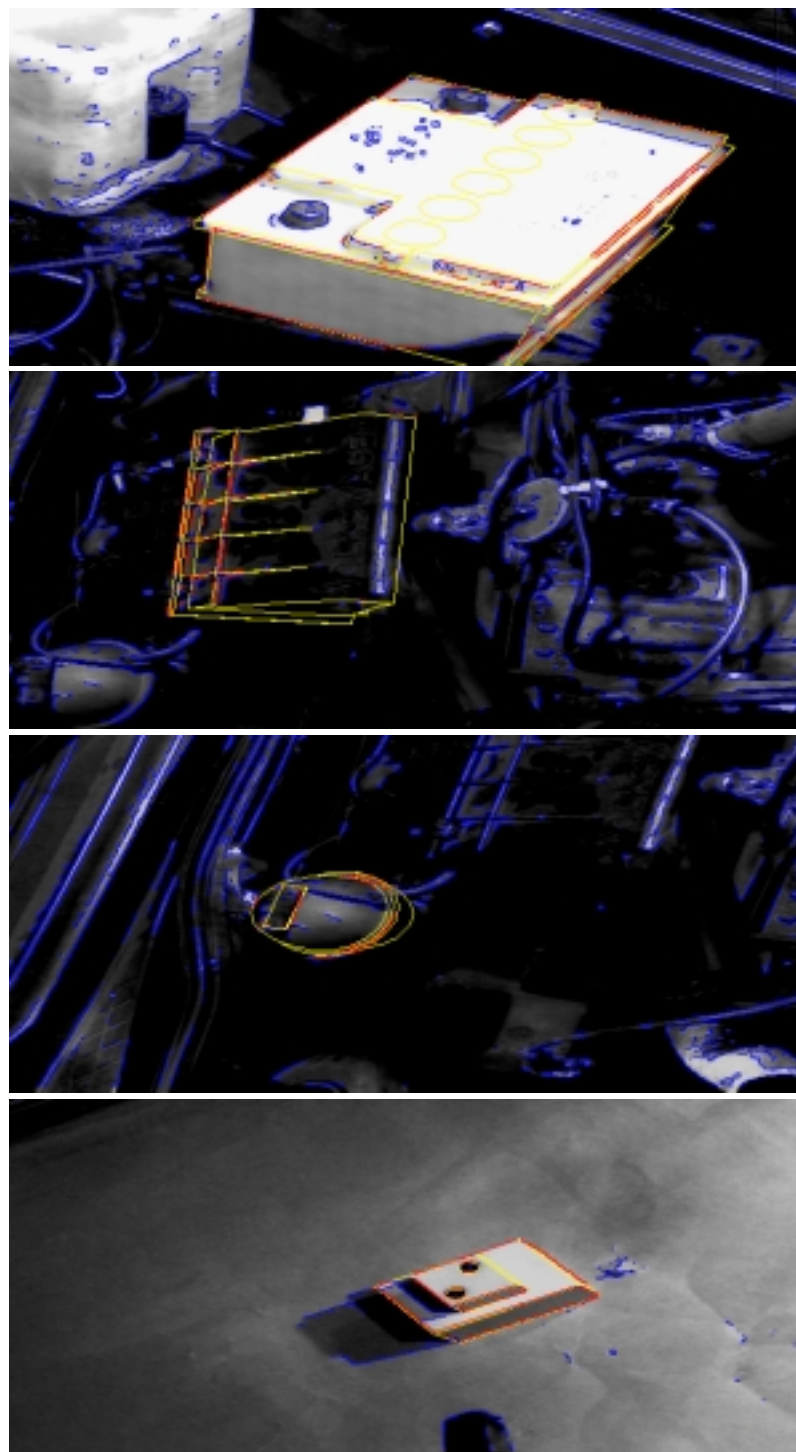


Abbildung 4.3: Monokulare Lokalisation der Batterie (oben), des Luftfilters (zweites Bild von oben), des Unterdruckbehälters (drittes Bild) sowie des Doppelklotzes (unten).

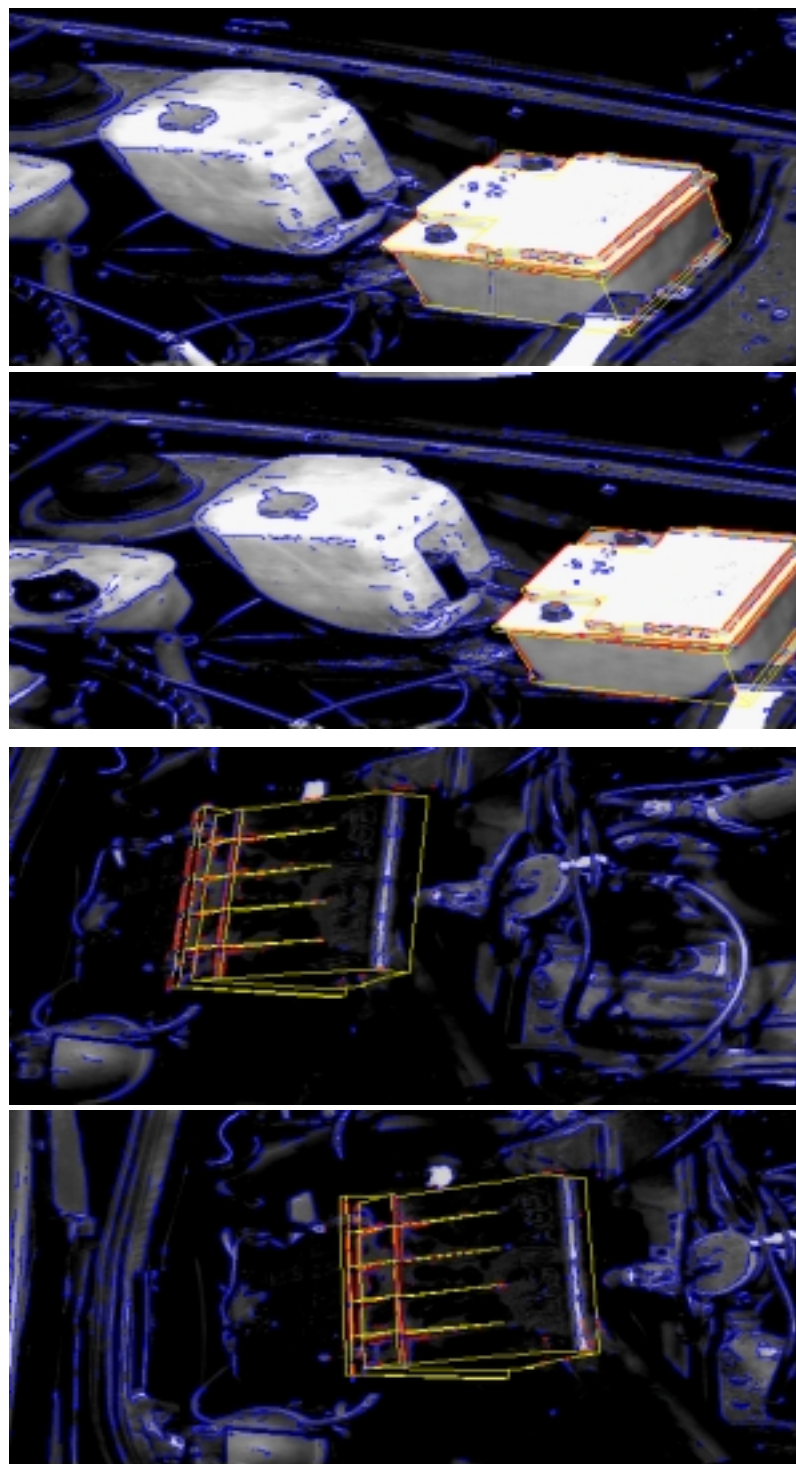


Abbildung 4.4: Binokulare Lokalisation der Batterie (oben: Projektion des Modells ins linke Kamerabild, darunter Projektion ins rechte) sowie des Luftfilters (drittes Bild von oben: Projektion ins linke Kamerabild, unten: Projektion ins rechte).

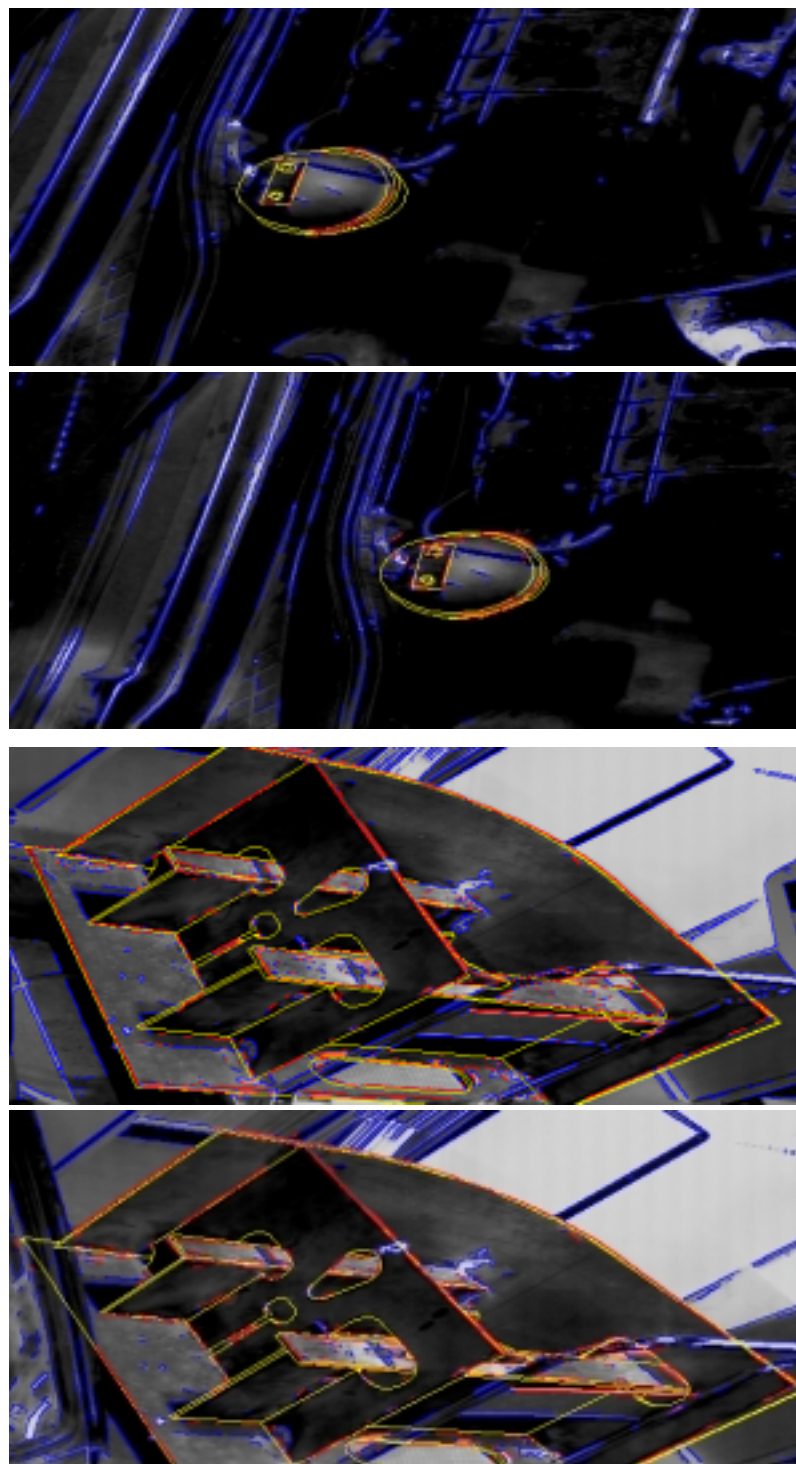


Abbildung 4.5: Binokulare Lokalisation des Unterdruckbehälters (oben: Projektion des Modells ins linke Kamerabild, darunter Projektion ins rechte) sowie des Schiffsteils (drittes Bild von oben: Projektion ins linke Kamerabild, unten: Projektion ins rechte).

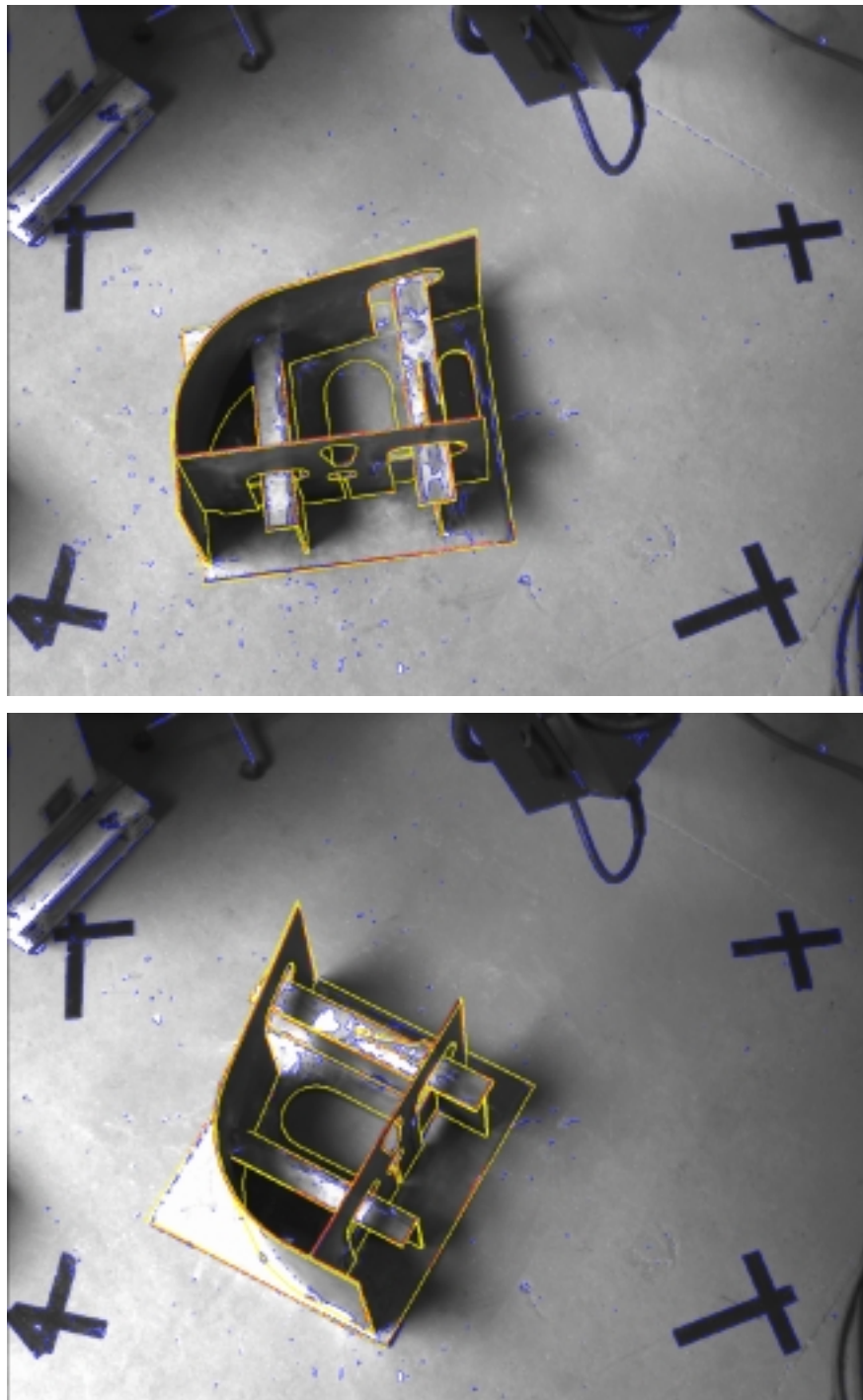


Abbildung 4.6: Ergebnisse der Lokalisationen des Schiffsteils in den hochauflösenden Bildern 1 und 2 (Die Liniendicke des überlagerten Modells wurde jeweils zur besseren Sichtbarkeit etwas vergrößert.)

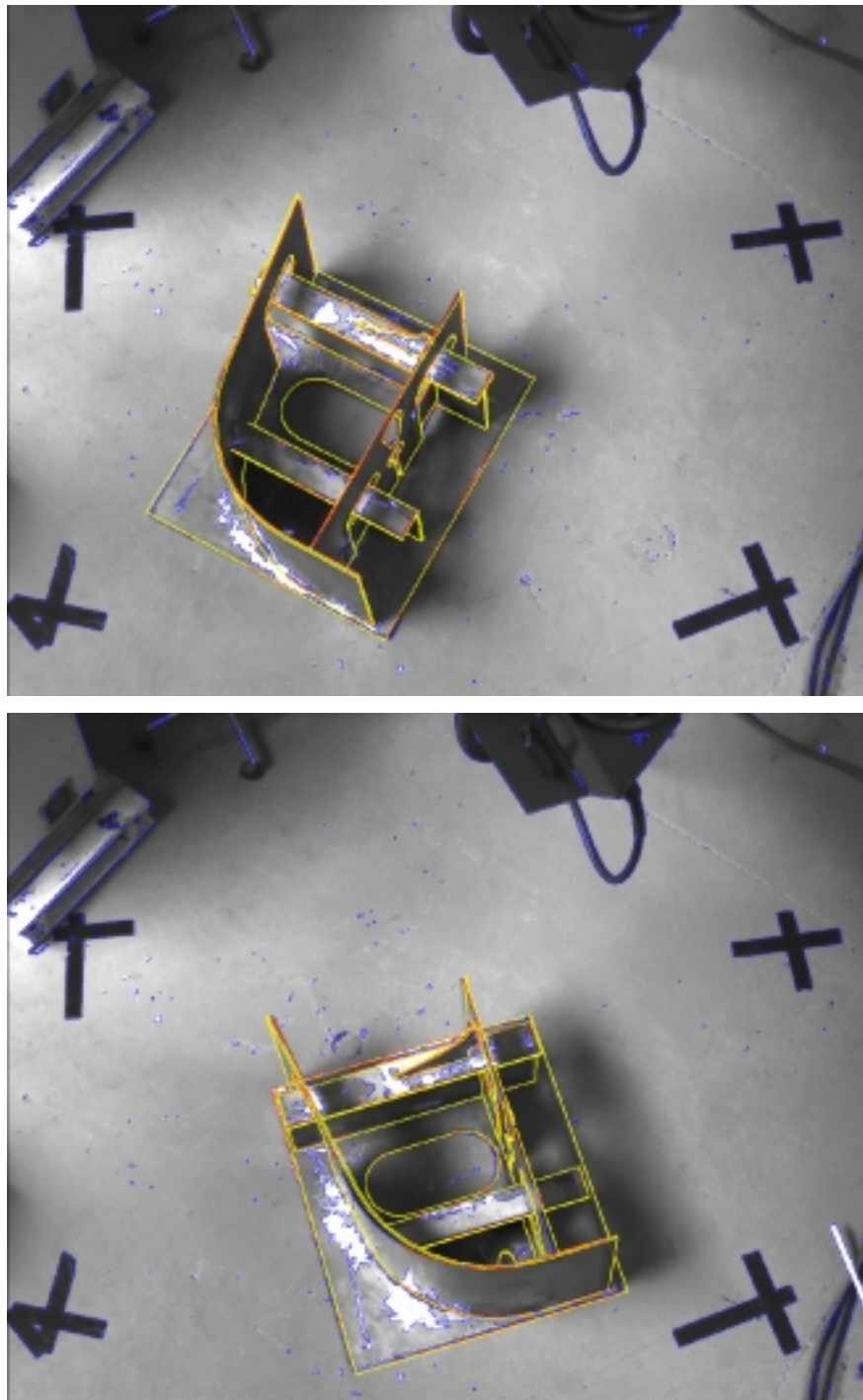


Abbildung 4.7: Ergebnisse der Lokalisationen des Schiffsteils in den hochauflösenden Bildern 3 und 4 (Die Liniendicke des überlagerten Modells wurde jeweils zur besseren Sichtbarkeit etwas vergrößert.)

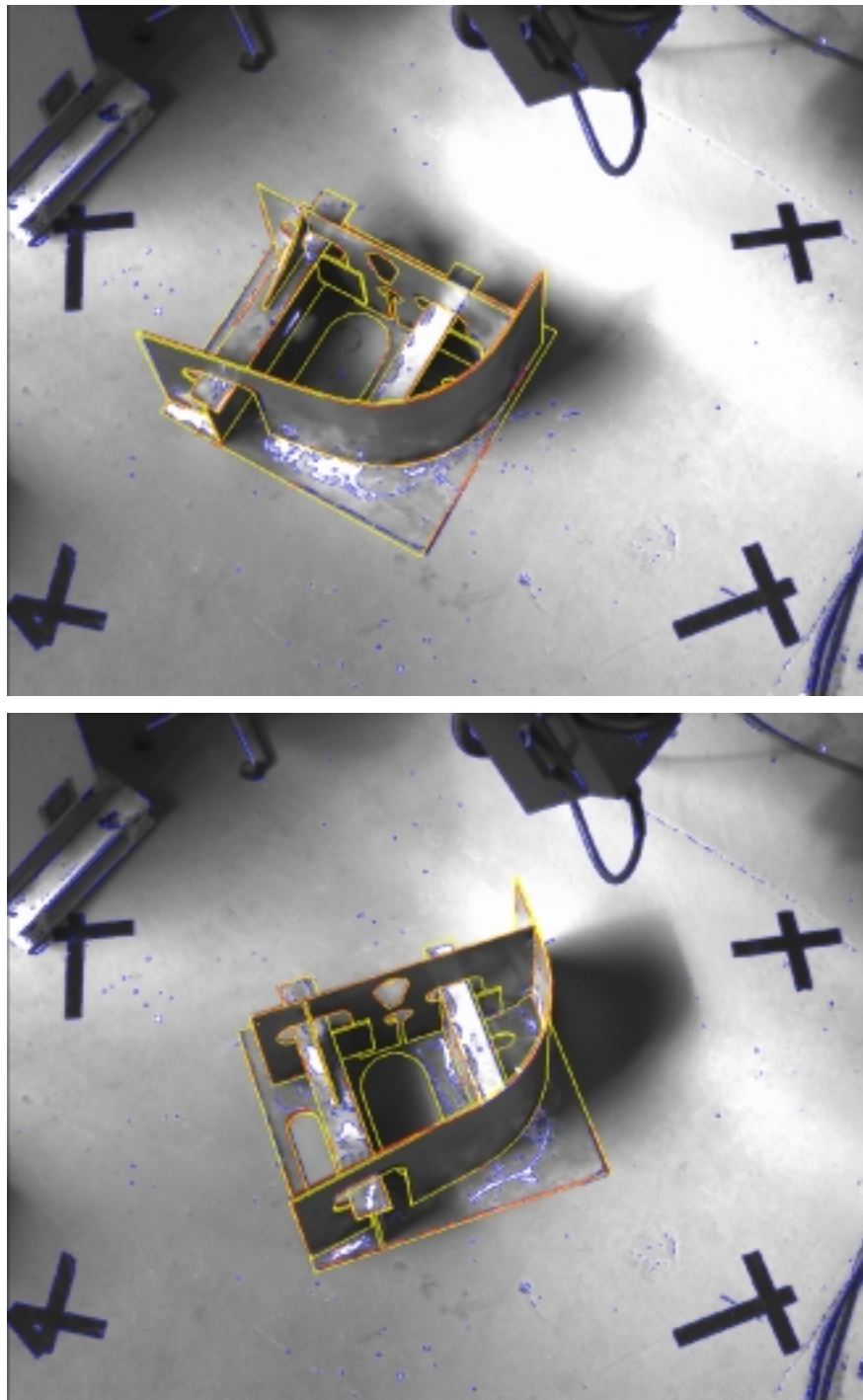


Abbildung 4.8: Ergebnisse der Lokalisationen des Schiffsteils in den hochauflösenden Bildern 5 und 6 (Die Liniendicke des überlagerten Modells wurde jeweils zur besseren Sichtbarkeit etwas vergrößert.)

4.2 Robustheit und Genauigkeit

Die in Abschnitt 3.1 eingeführten Definitionen für die Genauigkeit und Robustheit einer Lokalisation oder Verfolgung übertragen sich direkt auf die sichtsystemgestützte automatische Demontage. Eine Lokalisation bzw. Verfolgung gelte als robust bzw. genau, wenn $\Theta \leq \Theta_S := 3$ Pixel gilt.

Zusätzlich zur Robustheit bzw. Genauigkeit der Lokalisation und Verfolgung sind bei der Demontage auch die Begriffe der *Genauigkeit und Robustheit der Regelung* festzulegen. Wie bei den Begriffen zur Lokalisation und Verfolgung hängt auch hier die Robustheit von der erzielten Genauigkeit bei der Regelung ab und der Verschiedenartigkeit der durchgeführten Prozesse. Daher ist es wichtig, zum Nachweis der Robustheit möglichst unterschiedliche Demontagevorgänge unter möglichst unterschiedlichen Bedingungen zu testen. Die Genauigkeit ist in der realen Szene des Motorraumes zu messen, d.h. der Fehler zwischen gewünschtem Anfahrpunkt und tatsächlich angefahrenem ist zu bestimmen. Analog zur Genauigkeit bzw. Robustheit einer Lokalisation oder Verfolgung muß auch hier eine Schwelle – in dieser Arbeit Θ_R genannt – festgelegt werden, mit deren Hilfe entschieden werden kann, ob eine durchgeführte Regelung als genau bzw. robust einzustufen ist. Damit beispielsweise die automatische Entnahme der Batterie aus dem Motorraum erfolgreich durchgeführt werden kann, muß die Anfahrngenauigkeit des hierfür eingesetzten Greifers mindestens $\Theta_R := 0,5$ cm betragen. Aufgrund der Abmessungen von Greifer und Batterie setzt ansonsten eine der Greiferbacken während der Anfahrbewegung auf der Batterie auf, anstatt daß das Werkzeug die Greifposition bezüglich der Batterie erreicht, bei der sich die Batterie zwischen den Backen befindet. Bei einer Regelung zum automatischen Lösen von Schrauben mittels eines Schraubwerkzeuges muß der Schraubenschlitz mit einer Genauigkeit Θ_R von weniger als der halben Schlitzlänge angefahren werden, damit der Schrauber fassen kann.

Damit der Algorithmus zur sichtsystemgestützten automatischen Demontage Sinn macht, ist natürlich die Genauigkeit der eingesetzten Verfolgung Vorbedingung einer genauen Regelung. Das überträgt sich in natürlicher Weise selbstverständlich auch auf den Aspekt der Robustheit. Aus dieser Beziehung folgt, daß im Falle einer ungenauen Regelung zwischen folgenden beiden Phänomenen zu unterscheiden ist:

- Die Verfolgung ist bereits ungenau.
- Die Verfolgung ist genau, aber die Regelung ist es nicht.

In der vorliegenden Arbeit stellt sich heraus, daß bei allen hier dokumentierten Regelungsschwierigkeiten bis auf einen Fall⁶ eine ungenaue Verfolgung zugrunde liegt.

Der Sicherung der Robustheit bei der Verfolgung kommt daher bei der sichtsystemgestützten automatischen Demontage eine besonders große Bedeutung zu, da Roboter im Einsatz sind, die als Folge ungenauer Regelungen beschädigt werden können. Dieser

⁶Siehe den vierten Punkt der Liste von Abschnitt 4.2.2 auf Seite 123.

Aspekt bedeutet insbesondere auch eine entsprechend *angepaßte Arbeitsweise während der Programmentwicklung*. So können beispielsweise schon während der Entwicklung viele Risiken vermieden werden, indem bei abgeschalteten Robotern möglichst viel mit Bildern von Platte gearbeitet und getestet wird. Solche bei den Experimenten konstante Bildfolgen ermöglichen zudem numerische Vergleiche der ermittelten Lageschätzungen bei unterschiedlichen Versuchen und damit die frühzeitige Erkennung von eventuell bei der Entwicklung eingebauten Programmfehlern.

Im Gegensatz zur Personenverfolgung wird bei der Demontage aus Geschwindigkeitsgründen kein iterativ erweiterter Kalman-Filter eingesetzt. Tatsächlich wurde im Rahmen der Entwicklung der Arbeiten zu [Tonko 97] ein ursprünglich eingesetzter iterativ erweiterter Kalman-Filter mit Levenberg-Marquardt-Optimierung durch einen erweiterten Kalman-Filter ersetzt, um einer Echtzeitregelung so nahe wie möglich zu kommen. Da sich kein deutliches Indiz auf Mängel durch diese Optimierung ergeben, wird aufgrund der Priorität des Geschwindigkeitsaspekts bei der Regelung das Optimierungsverfahren des Aktualisierungsschrittes beibehalten. Die folgenden Abschnitte zeigen, daß die gegenwärtigen Robustheitsschwierigkeiten bei der Regelung an anderer Stelle zu suchen sind als bei der Personenverfolgung.

4.2.1 Verwendung einer adaptiven Bildfilterung

In [Tonko 97] (Abschnitt 3.4, ab Seite 49) wird das gravierendste Problem bei der Verfolgung angesprochen. Bei benachbarten parallelen projizierten Modellkanten treten Anpassungsschwierigkeiten aufgrund von Anpassungen der Kantenelemente an falsche projizierte Modellkanten auf. Dies wurde bisher dadurch umgangen, daß benachbarte parallele projizierte Modellkanten vom Anpassungsverfahren ausgeschlossen wurden (vgl. Abb. 3.7 in [Tonko 97]). Im Programm findet sich der Schwellwert 5 Pixel, d.h. alle parallelen projizierten Modellkanten, die keinen größeren Abstand als 5 Pixel voneinander haben, werden bei der Anpassung ignoriert. Wie sich bei der Verfolgung zeigt, führt dies insbesondere zu einer Ausblendung langer, im Bild oft sehr gut ausgeprägter Kanten. Da damit einige für eine robuste Verfolgung wichtige Kanten ignoriert werden, ist zu erwarten, daß die Robustheit der Verfolgung deutlich gesteigert werden kann, wenn diese Kanten so weit wie möglich integriert werden. Um diese Kanten im Verfolgungsprozeß integrieren zu können, müssen insbesondere die zugehörigen parallelen Bildkanten in der Bildfilterung besser aufgelöst werden ohne dabei gleichzeitig die Rauscheinflüsse zu verstärken (was ein Absenken der Gradientenbetragsschwelle zur Berechnung des Kantenelementbildes bewirken würde, vgl. Seite 6). Es wird erwartet, daß sich dadurch die Schwelle für den Minimalabstand der projizierten Modellkanten ohne Robustheitsverlust senken oder gar eliminieren läßt.

Zur Realisierung der adaptiven Bildfilterung wird der in [Nagel & Gehrke 98] beschriebene Ansatz (im 2D-Fall) verwendet und im folgenden verbessert. Als Testbeispiel eignet sich eine Aufnahme der Autobatterie, da im rechten Bereich im Bild besonders

viele parallele Bildkanten unterschiedlichen Abstands auftreten, siehe Abbildung 4.9 (oben). Als Referenzfilterung für die nachfolgenden Untersuchungen dient die Bildfilterung nach [Otte 94], da hierbei eine Vielzahl an Filterungsartefakten ausgeschlossen wird und sich diese bereits in anderen Anwendungen der Bildauswertung als sehr geeignet herausgestellt hat (siehe z.B. [Haag 98]). Der untere Teil von Abbildung 4.9 zeigt das nach [Otte 94] berechnete Kantenelementbild der Batterie.



Abbildung 4.9: Bild der Autobatterie (oben) und extrahierte Kantelemente bei Verwendung der Maskenberechnung mittels Abtastung einer abgeleiteten Gaußfunktion und der Maskenberechnung mit Randausgleich nach [Otte 94] (unten) bei einer Filterung mit konstanten Masken.

Das Verfahren nach [Nagel & Gehrke 98] ermöglicht es, bei der Berechnung der Masken für eine Tiefpaßfilterung bzw. Ableitung eines Grauwertbildes dessen lokale Gegebenheiten zu berücksichtigen. Dies geschieht, indem die Filtermasken lokal den Bildstrukturen entsprechend bestimmt werden. Dazu wird ein lokal bestimmter Grauwertstruktur-Tensor $GLST_{\text{init}}(\mathbf{x})$ zur Bestimmung der lokalen Filtermasken verwendet. Damit Filtermasken berechnet werden, die zu praktikablen Ergebnissen führen, müssen die Eigenwerte $\lambda_{\text{init},i}(\mathbf{x})$, $i = 1, 2$, von $GLST_{\text{init}}(\mathbf{x})$ in Anlehnung an [Nagel & Gehrke 98]

unter Beibehaltung der Eigenvektoren zuvor geeignet transformiert werden:

$$\lambda'_{\text{init},i}(\mathbf{x}) := \sigma_{\text{minsize}}^2 + \frac{\sigma_{\text{maxsize}}^2}{1 + \sigma_{\text{maxsize}}^2 \lambda_i^w(\mathbf{x})}, \quad (4.4)$$

wobei

$$\lambda_i^w(\mathbf{x}) := \lambda_{\text{init},i}(\mathbf{x}), \quad i = 1, 2. \quad (4.5)$$

Die Parameter $\sigma_{\text{minsize}}^2$ und $\sigma_{\text{maxsize}}^2$ geben dabei den Bereich für die resultierenden Kovarianzen an. In [Nagel & Gehrke 98] wird darauf hingewiesen, daß nur dann brauchbare Ergebnisse zu erwarten sind, wenn die Eigenwerte $\lambda_{\text{init},i}(\mathbf{x})$ zwischen $\sigma_{\text{minsize}}^2$ und $\sigma_{\text{maxsize}}^2$ liegen. Daher werden die Eigenwerte $\lambda_{\text{init},i}(\mathbf{x})$ gemäß [Nagel & Gehrke 98] geeignet normiert, d.h.

$$\lambda_i^w(\mathbf{x}) := \frac{\lambda_{\text{init},i}(\mathbf{x})}{\frac{1}{2} \text{Spur}(GLST_{\text{init}}(\mathbf{x}))}, \quad i = 1, 2 \quad (4.6)$$

anstatt der Festlegung nach (4.5) verwendet. Im Ausnahmefall $\text{Spur}(GLST_{\text{init}}(\mathbf{x})) = 0$ wird sinnvollerweise $\lambda_i^w(\mathbf{x}) := \lambda_{\text{init},i}(\mathbf{x}) := 0$ gesetzt. Abbildung 4.10 zeigt die sich ergebende adaptive Bildfilterung für $\sigma_{\text{minsize}}^2 = 0,5$ und $\sigma_{\text{maxsize}}^2 = 4,5$ bei einer Maskengröße von 7×7 . Diese Parameter werden auch im Folgenden verwendet – sofern nichts Anderes angegeben. Insbesondere bei den parallelen Bildkanten im rechten Teil der Batterie kann man gegenüber der Referenzfilterung aus Abbildung 4.9 deutliche Verbesserungen erkennen.



Abbildung 4.10: Adaptive Bildfilterung mit Eigenwerttransformation nach den Gleichungen (4.4) und (4.6). Im rechten Teil der Batterie kann man gegenüber dem unteren Teil von Abbildung 4.9 eine deutlich bessere Auflösung der parallelen Bildkanten erkennen.

Intuitiv zugänglicher wird die Wirkung der Eigenwerte auf die berechneten Masken, wenn die Transformation nach (4.4) und (4.6) ersetzt wird durch

$$\lambda'_{\text{init},i}(\mathbf{x}) := \lambda_i^w \sigma_{\text{minsize}}^2 + (1 - \lambda_i^w) \sigma_{\text{maxsize}}^2 \quad (4.7)$$

mit

$$\lambda_i^w(\mathbf{x}) := \frac{\lambda_{\text{init},i}(\mathbf{x})}{\text{Spur}(GLST_{\text{init}}(\mathbf{x}))} . \quad (4.8)$$

Die sich ergebenden extrahierten Kanten bei Verwendung dieser Transformation zeigt Abbildung 4.11. Im Vergleich zu Abbildung 4.10 kann man eine bessere Detektion der nahe beisammen liegenden, parallelen Kanten am rechten Teil der Autobatterie erkennen. Dies spricht eindeutig für die Verwendung von (4.7) und (4.8) anstatt (4.4) und (4.6).



Abbildung 4.11: Adaptive Bildfilterung mit Eigenwerttransformation nach den Gleichungen (4.7) und (4.8). Insbesondere die parallelen Bildkanten im rechten Teil der Batterie werden im Vergleich zu Abbildung 4.10 deutlich besser aufgelöst.

Es stellt sich die Frage, wie die Berechnung nach (4.7) und (4.8) weiter verbessert werden kann, d.h. wie $\lambda_i^w(\mathbf{x})$ und $\lambda'_{\text{init},i}(\mathbf{x})$ gewählt werden müssen, damit sich eine bezüglich der Modellanpassung optimale Kantenelementextraktion ergibt. Die Berechnung nach (4.7) und (4.8) führt zu umso kleineren Kovarianzen $\lambda'_{\text{init},i}(\mathbf{x})$, je größer der Eigenwert $\lambda_{\text{init},i}(\mathbf{x})$ gegenüber dem anderen Eigenwert $\lambda_{\text{init},3-i}(\mathbf{x})$ von $GLST_{\text{init}}(\mathbf{x})$ ist ($i = 1, 2$). Sind die beiden Eigenwerte von $GLST_{\text{init}}(\mathbf{x})$ gleich groß, so ergibt sich stets die gleiche Kovarianz $\lambda'_{\text{init},i}(\mathbf{x}) = \frac{1}{2}(\sigma_{\text{minsize}}^2 + \sigma_{\text{maxsize}}^2)$, $i = 1, 2$, und damit ergeben sich stets die gleichen adaptiven Masken zur Bestimmung der Ableitung am entsprechenden Bildpunkt. Insbesondere bedeutet dies im Falle gleich großer Eigenwerte, daß sich bei einer sehr schwachen Grauwertvariation die gleiche Filtermaske ergibt wie bei einer starken Grauwertvariation. Bei starken Grauwertvariationen wären allerdings im Vergleich zu geringen Grauwertvariationen kleinere Kovarianzen wünschenswert, damit in diesen Fällen die Grauwertstruktur besser lokalisiert werden kann, während bei schwachen Variationen eine stärkere Tiefpaßfilterung durchgeführt werden sollte. Um dies zu erreichen, kann man beispielsweise

$$\lambda'_{\text{init},i}(\mathbf{x}) := \left(\Delta_i \sigma_{\text{minsize}} + (1 - \Delta_i) \sigma_{\text{maxsize}} \right)^2 , \quad (4.9)$$

$$\Delta_i := \delta(\sqrt{\lambda_{\text{init},i}(\mathbf{x})}) \quad (4.10)$$

wählen mit der stückweise linearen Sigmoidfunktion

$$\delta : [0, 255] \rightarrow [0, 1], \quad \delta(x) := \begin{cases} 0, & \text{für } x < x_{\min} \\ \frac{x - x_{\min}}{x_{\max} - x_{\min}}, & \text{für } x_{\min} \leq x \leq x_{\max} \\ 1, & \text{für } x > x_{\max} \end{cases} \quad (4.11)$$

und den Parametern $x_{\min}, x_{\max} \in [0, 255]$, $x_{\min} < x_{\max}$, welche ein geeignetes Grauwertintervall für den linearen Anstieg von δ definieren. Zu bemerken ist, daß der Zusammenhang zwischen der Höhe $\sqrt{\lambda_{\text{init},i}(\mathbf{x})}$ eines Grauwertüberganges und der berechneten Standardabweichung $\sqrt{\lambda'_{\text{init},i}(\mathbf{x})}$ im Bereich $[x_{\min}, x_{\max}]$ linear ist.⁷

Durch Variation der Parameter x_{\min} und x_{\max} muß nun die bestmögliche Eigenwerttransformation ermittelt werden. Dahingehende Versuche zeigen, daß das beste Resultat für $x_{\min} = 10$ und $x_{\max} = 11$ erzielt wird. Wie man Abbildung 4.12 entnehmen kann, ist das erzielte Ergebnis deutlich besser als das in Abbildung 4.11 dargestellte. Wie weitere Experimente zeigen, können mit anderen Werten höchstens gleich gute Resultate erzielt werden, aber keine merklich besseren. Bei den gefundenen Werten für die Parameter x_{\min} und x_{\max} fällt auf, daß deren Differenz eins ist, d.h. die Sigmoidfunktion praktisch zu einer Sprungfunktion degeneriert. Es erstaunt, daß bei den transformierten Eigenwerten kein breiter Übergang zwischen $\sigma_{\text{minsize}}^2$ und $\sigma_{\text{maxsize}}^2$ benötigt wird, vgl. Gleichung (4.9).



Abbildung 4.12: Adaptive Bildfilterung mit Eigenwerttransformation nach den Gleichungen (4.9) und (4.10) für $x_{\min} = 10$, $x_{\max} = 11$. Insbesondere die parallelen Bildkanten im rechten Teil der Batterie werden im Vergleich zu Abbildung 4.11 deutlich besser aufgelöst.

Die Frage nach der besten Eigenwerttransformation reduziert sich demnach auf die Suche nach einer Schwelle, bis zu der die Höhe einer Grauwertkante vernachlässigt, d.h.

⁷Es ist darauf hinzuweisen, daß aufgrund der Definition des lokalen Grauwertstruktur-Tensors dessen Eigenwerte $\lambda_{\text{init},i}(\mathbf{x})$ proportional zum *Quadrat der Höhe des Grauwertüberganges* sind.

implizit dem Rauschen oder unwesentlichen Grauwertvariationen zugeschrieben wird, und ab welcher Grauwertkantenhöhe ein Pixel als Kantenelement, d.h. als struktureller Grauwertübergang angesehen werden soll. Die Suche nach einer solchen Schwelle ist äquivalent zu der Suche nach einer geeigneten Funktion, welche die Schwelle ersetzen soll, d.h. – unabhängig davon, ob eine Schwelle oder eine Funktion gesucht wird – in beiden Fällen muß festgelegt werden (im einen Fall explizit, im anderen implizit), wo die Grenze zwischen unwesentlicher/stochastischer und struktureller Grauwertvariation gezogen werden soll. Obige Untersuchungen belegen, daß offenbar diese Grenze scharf sein darf, also kein breiter kontinuierlicher Übergang erforderlich ist. Erwartungsgemäß variiert diese Grenze je nach zu filterndem Bild. Im vorliegenden Fall ist 10 ein guter Wert. Eine unscharfe Grenze, d.h. ein breiterer Bereich $[x_{\min}, x_{\max}]$ (evtl. mit einer anderen, nicht stückweise linearen Sigmoidfunktion δ) könnte hier möglicherweise genutzt werden, um bei vielen möglichen, unterschiedlichen Eingabebildfolgen einen Kompromiß für die Schwelle zu finden, d.h. die Transformation für den Fall allgemeiner Bildfolgen zu optimieren.

Man mag nun argumentieren, daß die kleinere Kovarianz der Filtermaske maßgeblich für die erzielten Resultate verantwortlich ist und die größere keine Rolle spielt. Dies kann sofort widerlegt werden mit einer Filterung für $x_{\min} = -1$ und $x_{\max} = 0$, d.h. bei konstanter Kovarianz $\lambda'_{\text{init},i}(\mathbf{x}) \equiv \sigma_{\text{minsize}}^2$. Das Filterergebnis ist wesentlich schlechter als alle hier dargestellten Ergebnisse und schlechter als die Referenzfilterung aus Abbildung 4.9. Dies belegt, daß nicht nur die geschickte Wahl der Kovarianz für die Richtung senkrecht zu den Grauwertkanten im Hinblick auf eine gute Kantendetektion wichtig ist, sondern daß auch die Kovarianz in Richtung der Grauwertkanten eine entscheidende Rolle spielt. Damit ist nachgewiesen, daß durch eine große Kovarianz tangential zu einer Grauwertübergangsfrent die (aufgrund einer kleinen Kovarianz) schwache Tiefpaßfilterung senkrecht zur Grauwertübergangsfrent ausgeglichen wird.

Aufgrund der damit nachgewiesenen Wichtigkeit *beider* berechneter Kovarianzen $\lambda'_{\text{init},i}(\mathbf{x})$, $i = 1, 2$ bietet es sich an, bei vorliegenden Grauwertkanten deren Differenz zu vergrößern. Damit wird möglicherweise eine zusätzliche Verstärkung schwacher Grauwertkanten bewirkt. Um diese Hypothese zu testen, erfolgt versuchsweise die Eigenwerttransformation nach (4.9) unter Verwendung von

$$\Delta_i := \delta \left(\underbrace{\frac{\lambda_{\text{init},i}(\mathbf{x})}{\text{Spur}(GLST_{\text{init}}(\mathbf{x}))}}_{\in[0,1]} \underbrace{\sqrt{\lambda_{\text{init},i}(\mathbf{x})}}_{\in[0,255]} \right), \quad i = 1, 2 \quad (4.12)$$

bei $x_{\min} = 10$ und $x_{\max} = 11$. Das Ergebnis ist mit bloßem Auge nicht von Abbildung 4.12 zu unterscheiden.

Eine andere Möglichkeit, die Differenz zwischen den berechneten Kovarianzen $\lambda'_{\text{init},i}(\mathbf{x})$, $i = 1, 2$ zu vergrößern, ergibt sich aus folgender Überlegung: Für die Modellanpassung sind vor allem die extrahierten langen Bildkanten von Bedeutung. Die Enden extrahierter Kanten sind oft gegenüber dem eingezogenen Bild nicht korrekt lokalisiert. Wenn z.B. zwei Bildkanten in spitzem Winkel aufeinander treffen, beeinflußt

die eine Bildkante den detektierten Verlauf der anderen desto mehr, je näher man sich beim Schnittpunkt der Bildkanten befindet. An solchen Schnittpunkten ergeben sich zwei große Eigenwerte $\lambda_{\text{init},i}(\mathbf{x})$ ($i = 1, 2$). Für die Modellanpassung sind vor allem die Stellen wichtig, an denen ein Eigenwert groß und der andere klein ist. Es ist daher zu erwarten, daß es vorteilhaft ist, solche Bereiche auszudehnen. Um das zu testen, wird versuchsweise $\lambda'_{\text{init},i}(\mathbf{x})$ wie in (4.9) gewählt unter Verwendung von

$$\Delta_i := \delta(\sqrt{\lambda_{\text{init},i}(\mathbf{x})} \delta_{i,3-i}) , \quad i = 1, 2 \quad (4.13)$$

und

$$\delta_{j,k} := \begin{cases} 0 & \text{für } \lambda_{\text{init},j}(\mathbf{x}) \leq \lambda_{\text{init},k}(\mathbf{x}) \\ 1 & \text{für } \lambda_{\text{init},j}(\mathbf{x}) > \lambda_{\text{init},k}(\mathbf{x}) \end{cases} \quad (j, k \in \{1, 2\}).$$

Auch das hiermit erzielte Ergebnis für $x_{\min} = 10$, $x_{\max} = 11$ ist mit bloßem Auge nicht von Abbildung 4.12 zu unterscheiden. Offenbar lassen sich die Ergebnisse, die mittels der Transformation (4.9), (4.10) in Abbildung 4.12 erzielt wurden, durch Verfahrensmodifikationen zumindest nicht mehr signifikant verbessern. Versuche zur Optimierung der Werte für $\sigma_{\text{minsize}}^2$ und $\sigma_{\text{maxsize}}^2$ führen ebenfalls zu keinen besseren Resultaten als bisher. Daher wird die Suche nach einer optimalen Transformation bzw. Parametrisierung als abgeschlossen betrachtet.

Die Verfolgung der Autobatterie unter Verwendung der adaptiven Bildfilterung funktioniert so genau wie die bisherige Bildfilterung bei einer Schwelle von 5 Pixeln für den Minimalabstand der projizierten Modellkanten. Um den Robustheitsgewinn durch die Einführung der adaptiven Bildfilterung zu ermitteln, wird schrittweise die Schwelle für den Minimalabstand der projizierten Modellkanten erniedrigt. Die Schwelle kann bis auf 2,75 Pixel herabgesetzt werden, ohne daß Robustheitseinbußen bei der Verfolgung mit adaptiver Bildfilterung beobachtet werden, siehe Abbildung 4.13. Bei der Filterung mit ortsinvarianten Filtermasken treten dagegen bei dieser Schwelle schlechte Anpassungen auf.

Zusammenfassend ist zu sagen:

- Parallele Bildkanten werden durch die Integration und Verbesserung der adaptiven Bildfilterung gemäß der Erwartungen deutlich besser herauspräpariert, ohne gleichzeitig Rauscheinflüsse zu verstärken.
- Dadurch wird die Robustheit der Verfolgung deutlich gesteigert.
- Insbesondere kann die Schwelle für den Minimalabstand projizierter Modellkanten im Bild signifikant abgesenkt werden, wodurch mehr Modellkanten im Anpassungsprozeß berücksichtigt werden.



Abbildung 4.13: Verfolgung der Autobatterie zum Halbbildzeitpunkt 201 bei der bisherigen Bildfilterung mit ortsinvarianten Filtermasken (oben) und der adaptiven Bildfilterung nach (4.9), (4.10) mit $x_{\min} = 10$, $x_{\max} = 11$ (unten) bei einer Schwelle für den Minimalabstand projizierter Modellkanten von 2,75 Pixel.

4.2.2 Verbesserung verschiedener Einzelfaktoren

Wie sich bereits bei der Personenverfolgung gezeigt hat, ist die Robustheitssteigerung mit der Analyse und Verbesserung zahlreicher Einzelfaktoren verbunden. Dies bestätigt sich auch bei der sichtsystemgestützten automatischen Demontage. Auch hier liegt oft nicht in den Maßnahmen zur Verbesserung selbst die eigentliche Schwierigkeit (einige Maßnahmen sind sehr einfach bei gleichzeitig großer Wirkung), sondern in der Analyse der zugrundeliegenden Ursachen. Durch systematische Rückverfolgung sich zeigender Robustheitsschwächen werden hier deutliche Verbesserungen erzielt. Insgesamt tragen folgende Maßnahmen zur Steigerung der Robustheit bei:

- Verbesserung der Einstellung der Kovarianzen (in [Tonko 97] mit P_0^- und Q bezeichnet) für die Unsicherheit der Lageschätzung der Modelle. Als sehr gute Richtlinie hat sich dabei bewährt, die maximale Differenz der während einer genauen Verfolgung berechneten Zustandsvektoren für das Systemrauschen Q anzusetzen. Insbesondere die Robustheit der Stereoverfolgung von Batterie und Greifer wird dadurch gegenüber [Tonko 97] deutlich verbessert.

- Beseitigung eines Relikts aus der Zeit des einst verwendeten iterativ erweiterten Kalman-Filters mit Levenberg-Marquardt-Optimierung (siehe Seite 115): Der Levenberg-Marquardt-Parameter μ zur Interpolation zwischen Gauß-Newton- und Gradientenabstiegsverfahren (zur Bedeutung von μ vgl. Algorithmus von Seite 47) wird fälschlicherweise mit einem nicht zu vernachlässigenden Wert von 0,1 weiterhin verwendet und vor der Invertierung der zweiten Ableitung der Gütefunktion eingerechnet.
- Durch die in Abschnitt 4.5.1 erläuterte Beschleunigung ergibt sich ebenfalls eine Robustheitssteigerung, da – bei unveränderter Bewegungsgeschwindigkeit des Werkzeugroboters – deutlich mehr Kamerabilder ausgewertet werden als in [Tonko 97].
- Der Regelungsmechanismus selbst wird verbessert. Insbesondere wird ein besseres Abstandsmaß zum Anfahren der Zielposition eingesetzt, welches die Verhältnisse in der Szene besser modelliert: Bisher wurde als Kriterium für das Erreichen der Zielposition die Änderung der Roboter-Hand-Transformation verwendet. Wenn diese Änderung unter eine Genauigkeitsschwelle fiel, wurde die Anfahrposition als erreicht angesehen. Nun wird die euklidische Abweichung zwischen gewünschter Zielposition und tatsächlicher Position verwendet und mit einer vorgegebenen Genauigkeitsschwelle verglichen. Dadurch kann das verwendete Werkzeug relativ zum Werkstück wesentlich genauer als bisher positioniert werden. Zudem können im Gegensatz zu vorher – wo kein direkter Zusammenhang zur tatsächlich erreichten Genauigkeit bestand – die vorgegebenen Anfahrpräzisionen bezüglich der Lageschätzungen der Objekte zugesichert werden.
- Während es sich in [Tonko 97] aus Rechenzeitgründen als sinnvoll herausgestellt hat, nur wenige Teile des Greiferwerkzeuges zu modellieren (siehe z.B. [Tonko 97], Seite 96; Modell dargestellt im linken Teil von Abbildung 4.14), so ergibt sich nach obigen Maßnahmen eine andere Situation bei der Greiferverfolgung. Einige auftretende Fehlanpassungen des Greifers können durch Erweiterung des Modells beseitigt werden, siehe Abbildung 4.14 (rechts).

Als Bewertungsmaßstab der angegebenen Arbeiten kann das Folgende dienen. Vor Durchführung der genannten Maßnahmen reagierte die automatische Entnahme der Batterie aus dem Motorraum so empfindlich auf Rahmenbedingungen wie z.B. leichte Variationen der Stärke des in der Roboterzelle einfallenden Tageslichts, daß die Regelung in vielen Fällen mißlang. Die sich dabei ändernden Rahmenbedingungen waren mit bloßem Auge oft nicht festzustellen. Meist setzte dabei eine der Greiferbacken beim Anfahren der Zielposition auf der Batterie auf oder der Greifer bewegte sich fälschlicherweise auf die Karosserie des Fahrzeugs zu und das Experiment mußte abgebrochen werden. Dagegen funktioniert nun die Entnahme der Batterie aus dem Motorraum bei praktisch allen Programmläufen robust. Lediglich bei starken Änderungen der verwendeten Beleuchtungsbedingungen treten noch Anpassungsprobleme des Greifers auf.

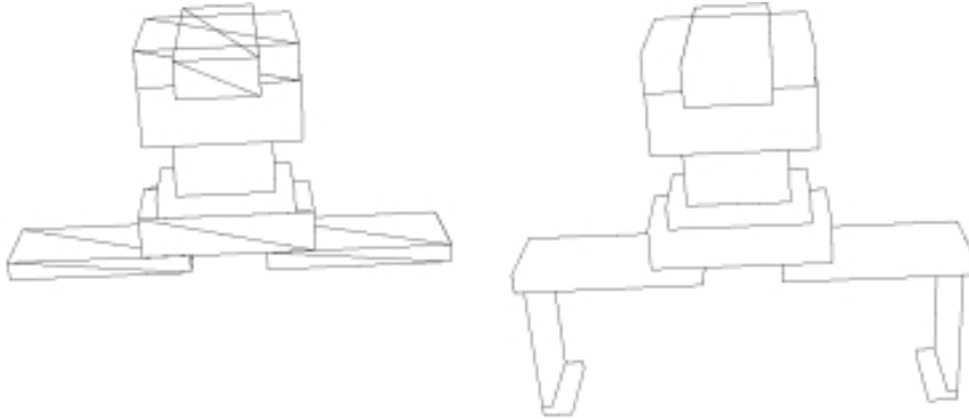


Abbildung 4.14: Modell des Greiferwerkzeuges vor (links) und nach (rechts) der Erweiterung. Die im linken Teil durchgestrichenen Objektteile waren zwar bei [Tonko 97] modelliert, wurden aber während der Verfolgung deaktiviert, um die Zykluszeit der Verfolgung hinreichend gering zu halten. Da die Greiferbacken aus Aluminium bestehen, sind einige Bildkanten sehr kontrastschwach. Dies wird während der Verfolgung berücksichtigt, indem entsprechende projizierte Modellkanten nicht eingezeichnet werden (siehe insbesondere rechte Teilabbildung).

Diese sind in allen beobachteten Fällen darauf zurückzuführen, daß Modellkanten des Greifers an Hintergrundkanten (Kanten des Kotflügels oder der Autobatterie) angepaßt werden, die hinsichtlich ihrer Position und Richtung kompatibel zu den Modellkanten sind.

Ein weiterer Beleg für die erzielte Robustheit ergibt sich aus der genauen Anpassung des Schiffsteils unter signifikanter partieller Verdeckung, siehe Abbildung 4.24 auf Seite 141.

4.3 Zuverlässigkeit, Sicherheit, Ausfallsicherheit

Während es in [Tonko 97] darum ging, überhaupt erst ein Verfahren zu entwickeln, das sightssystemgestützt Roboter regeln kann, mußten verschiedene andere Aspekte damals aus verständlichen Gründen zurückgestellt werden. Dazu zählt insbesondere auch die Zuverlässigkeit des Systems, die Sicherheit für die eingesetzten Geräte, die Ausfallsicherheit des Systems, sowie die Analysierbarkeit der Ergebnisse bzw. des Programms (siehe Abschnitt 4.6). Daher zeigen sich bei dem System in der Fassung nach [Tonko 97] während der Verfolgung bzw. Regelung verschiedene Störeffekte, die zu unvorhersehbaren Zeiten auftreten und sich in unterschiedlichster Art und Weise bemerkbar machen. Beispielsweise versagt der Bildeinzug gelegentlich unter Ausgabe unterschiedlicher Meldungen seinen Dienst und die Verfolgung bzw. Regelung wird abgebrochen, einer der Roboter ist bei einem Programmstart oder plötzlich während der Regelung nicht (mehr) ansprechbar, das System stürzt in einigen Fällen scheinbar grundlos ab, der Greifer

wird trotz funktionierender Verfolgung an eine falsche Stelle gefahren oder die Verfolgung selbst versagt ohne erkennbaren Grund. Bei dem letzteren Punkt überträgt sich die Frage nach der Robustheit direkt auf den Sicherheitsaspekt. Die Verbesserung der Robustheit aus Abschnitt 4.2.2 bedeutet folglich gleichzeitig eine Verbesserung der Sicherheit für die eingesetzten Geräte und natürlich auch die Ausfallsicherheit. Da die übrigen Punkte in nicht reproduzierbarer, (scheinbar) zufälliger Art und Weise auftreten, ist eine Rückverfolgung der Ursachen besonders schwierig. Es ergibt sich ein System, das scheinbar zufällig immer neue Fehler produziert und manchmal zur korrekten Durchführung der Entnahme der Batterie aus dem Motorraum mehrmals gestartet werden muß. Zudem werden Steuerparameter zur Parametrisierung des Bildeinzugs und der Kantenelementberechnung mit z.T. falschen Werten an die MiniVISTAs (vgl. Seite 6) durchgereicht.

Beim Robotereinsatz ist es natürlich zum Schutz der Geräte besonders wichtig, solche Probleme zu beseitigen. Um möglichst viele Indeterminismen im Programm aufspüren zu können, erweist sich die Verwendung von mehr als einem Übersetzer (siehe Abschnitt 4.4) und der aktivierbaren Warnungsmeldungen beim Übersetzen des Programms als wertvolles Hilfsmittel (vgl. auch Abschnitt 3.3). Die Verwendung von mehr als einem Übersetzer hat dabei den Vorteil, daß ein Übersetzer gelegentlich mögliche Indeterminismen bzw. mögliche Fehlerquellen erkennen kann, die der andere nicht zu dokumentieren in der Lage ist.

Da die GIGAmachine selbst in seltenen Fällen Fehler produziert und die Herstellerfirma mittlerweile nicht mehr für eine Reparatur existiert, trägt der Umstieg auf einen anderen Rechner dazu bei, das Regelungssystem ausfallsicherer zu machen, vgl. Abschnitt 4.5.1.

Bei dem Versuch, die nun noch verbleibenden Störursachen aufzuspüren, stellt sich heraus, daß viele davon durch das Zusammenspiel komplexer und schwer analysierbarer Programmteile sowie einer unnötig fehleranfälligen Zusammenschaltung der eingesetzten Spezialrechner (Transputer) zustande kommen. Dies ist zunächst näher zu erläutern. Bei der Verschaltung der zur Ansteuerung der Roboter notwendigen Spezialrechner sowie der MiniVISTAs konnten zu Zeiten der GIGAmachine nur zwei Karten zur Ansteuerung der Geräte verwendet werden. Das bedeutet, daß die Ansteuerung der Roboter durch eine MiniVISTA 'hindurchgeschleift' werden mußte, siehe Abbildung 4.15 oder auch die Abbildung auf Seite 107 in [Tonko 97]. Durch die Konkurrenz von Bildeinzug und Roboteransteuerung auf der Verbindung zum Rechner läßt sich die Datenübertragung nur sehr schwer ausfallsicher realisieren, was zu entsprechend komplexen – und dennoch nicht hundertprozentig ausfallsicheren – programmtechnischen Ansteuerungen führen mußte. Zudem muß die sternförmige Verschaltung der nicht am Bildeinzug beteiligten Spezialrechner berücksichtigt werden. Dies wird realisiert durch generische Programmkonstrukte, die selbstständig die Topologie des verwendeten Rechnernetzes erkennen können und die Daten entsprechend verschicken. Um die Ansteuerungen zu unterstützen wurde ein Präprozessor für C/C++ [Schäfer 96] eingesetzt, der durch Generierung entsprechender Programmquelltexte die Arbeit erleichtern

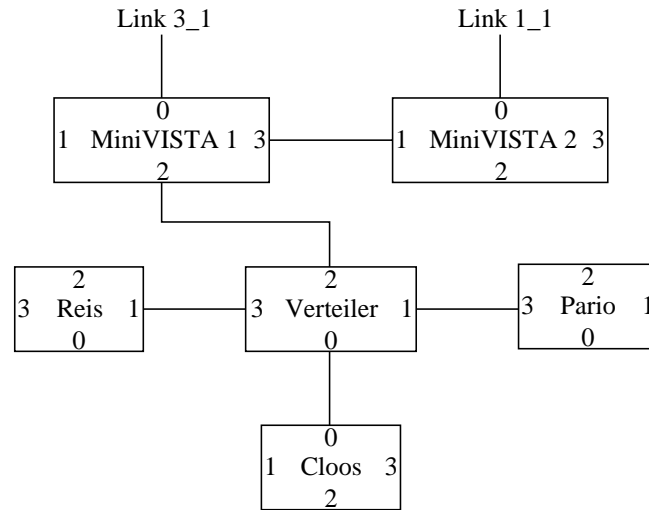


Abbildung 4.15: Verschaltung der Spezialrechner (Transputer) zur Geräteansteuerung. Die beteiligten Rechner sind durch Rechtecke dargestellt. Die Zahlen innerhalb der Rechtecke geben die Verbindungen ('Links') an, über welche die Rechner mit anderen in Kontakt treten können. MiniVISTA 1 ist zuständig für den Bildeinzug von der rechten Kamera des Beobachtungsroboters und MiniVISTA 2 für den der linken. Die Verbindung zwischen den MiniVISTAs dient der Stereobildsynchronisation. `/dev/link/3_1` und `/dev/link/1_1` bezeichnen die Eingänge am Steuerrechner. Reis bezeichnet den Rechner für den Beobachtungsroboter und Cloos den des Werkzeugroboters. Pario diente einst der Einstellung von Blenden- und Schärferegler der Kamera auf dem Endeffektor des Werkzeugroboters vom Rechner aus (wird aber jetzt nicht mehr verwendet).

sollte. Das Konzept, die Daten über evtl. mehrere Rechner zum Zielrechner hindurchzureichen wurde zudem kombiniert mit Mechanismen zum Fernstarten von Funktionen auf anderen Rechnern ('remote procedure calls'). Die Kombination all dieser Mechanismen verknüpft mit automatisch vom Präprozessor generiertem und daher sehr schwer lesbarem Programmquelltext erweist sich nun als so komplex, daß es fast unmöglich erscheint, die auftretenden Schwierigkeiten darin lokalisieren zu können.

Da sich besagte Programmstruktur wie auch die Spezialrechnerverschaltung nun als *unnötig* komplex herausstellt und nicht mehr zwingend benötigt wird, entscheidende Schwierigkeiten verursacht und zudem die Verbesserung des Systems schwierig oder sogar unmöglich macht, wird entschieden, im Hinblick auf Zuverlässigkeit, Ausfallsicherheit und Analysierbarkeit des Systems eine möglichst einfache Struktur zu wählen. Insbesondere werden die Spezialrechner angemessener verschaltet, siehe Abbildung 4.16. Während bisher bei Ausfall eines Spezialrechners ein 'Wurmprogramm' das Netz nach dem abgestürzten Gerät auskundschaften mußte, hat diese Maßnahme den zusätzlichen Vorteil, daß das Gerät direkt vom Steuerprogramm erkannt wird.

Zur Verbesserung von Zuverlässigkeit, Ausfallsicherheit und Analysierbarkeit werden systematisch u.a. folgende Bereinigungen am Programmsystem durchgeführt (für De-

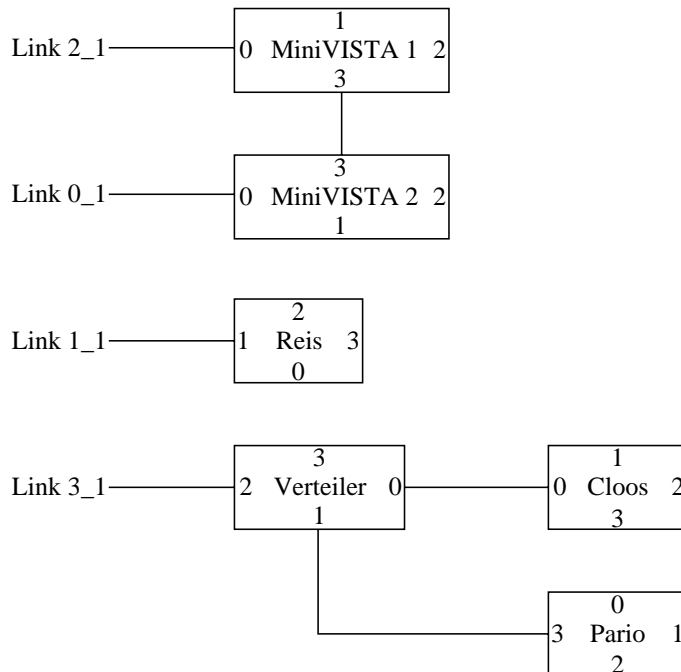


Abbildung 4.16: Verbesserte Verschaltung der Spezialrechner (Transputer) zur Geräteansteuerung. Zur Darstellung siehe Bildunterschrift zu Abb. 4.15. Es werden nun vier Eingänge am Steuerrechner verwendet (`/dev/link/0_1` bis `/dev/link/3_1`). **Pario** ist programmtechnisch stillgelegt und **Verteiler** reicht alle ankommenden Daten zwischen den Verbindungen 0 und 2 direkt durch (falls **Pario** in Zukunft wieder verwendet werden soll, kann man dies einfach ändern).

tails siehe [Müller & Nagel 99]):

- Elimination der oben genannten, nun unnötig gewordenen komplexen Konzepte,
- dazu Reimplementation von Bildeinzug und Roboteransteuerung sowie Verbesserung einiger Teile, vgl. Abschnitt 4.4.1,
- Beseitigung des Präprozessors zur Verbesserung der Lesbarkeit und Analysierbarkeit des Programms,
- weitere unnötig gewordene Programmteile werden ausfindig gemacht und eliminiert (durch Wegfallen des Präprozessors können weitere generische Konstrukte entfernt werden, z.B. bei der Roboteransteuerung; zudem finden sich viele Routinen, die auf der GIGAmachine notwendig waren und nun nach dem Umstieg auf den neuen Steuerrechner entfallen können, vgl. S. 142).
- Verbesserung der Analysierbarkeit von eventuellen Fehlern im Programm durch Verbesserung der Einbindestructur bei den Quelldateien sowie Vereinfachung und Verkürzung einiger Programmteile.

Durch diese Maßnahmen wird insbesondere auch die *Wartbarkeit* des Systems deutlich spürbar (zu spüren an den für Umbauten benötigten Zeiten) verbessert. Tabelle 4.2 zeigt, von welchem bedeutenden Umfang die unnötig gewordenen Programmteile vor der Bereinigung des Systems tatsächlich waren.

	vorher	nachher	Verbesserung
Programmgröße ohne Objektdateien	19,3 MB	2,8 MB	85,5%
Visdom-Programmtext	96 Dateien	43 Dateien	55,2%
	1,55 MB	1,08 MB	30,3%
	32501 Zeilen	24585 Zeilen	24,4%
Übersetzungsdauer (nicht optimiert)	213 s	51 s	76,1%
Beim Übersetzen erzeugte Dateien (nicht optimiert übersetzt)	108 Dateien	29 Dateien	73,1%
	5,7 MB	2,3 MB	59,6%
Länge der ausgeführten Datei	4,4 MB	0,72 MB	83,6%

Tabelle 4.2: Verbesserung des Visdom-Systems durch die Bereinigungsaktionen, quantifiziert anhand einiger Meßgrößen.

Auf funktioneller Ebene zeigt sich, daß nun die Parametrisierung des (verbesserten) Bildeinzugs korrekt durchgeführt wird. Zudem werden die eingangs genannten Störungen bei der Regelung nur noch beobachtet, wenn die Verbindung zu einem der Spezialrechner unterbrochen oder aufgrund des Alterungsprozesses oxidiert ist. Die Regelung des Greifers zur Batterieentnahme funktioniert bei praktisch jedem Programmlauf einwandfrei (bis auf den in Abschnitt 4.2.2 dokumentierten Sonderfall bei extremen Beleuchtungsänderungen). Die Bereinigungsaktionen haben also eine wesentliche Auswirkung auf die Ausfallsicherheit des Systems.

Die hier durchgeführten Untersuchungen belegen, daß unnötig gewordene Programmteile sowie unnötig komplexe oder umfangreiche Programmteile signifikante Einbußen bezüglich vieler Bewertungsaspekte, wie Zuverlässigkeit, Sicherheit, Ausfallsicherheit, Robustheit, Wartbarkeit und Analysierbarkeit nach sich ziehen können. Auf den allgemeinen Fall übertragen bedeutet dies, daß solche Programmteile *nicht nur* im verwendeten Programmsystem *brach* liegen, sondern *prinzipielle Risiken* in sich bergen. Die hier dokumentierten Auswirkungen des Grundsatzes „*Ein System sollte so komplex sein, wie es die zu erfüllende Aufgabe erfordert, aber nicht komplexer*“ sprechen für sich.

4.4 Flexibilität

Zur Erhöhung der Flexibilität des eingesetzten Verfolgungs- bzw. Regelungssystems spielt zum einen eine weitgehende Unabhängigkeit von den verwendeten Geräten eine wichtige Rolle, um das System bei Bedarf universell auch in anderen Einsatzfällen anwenden bzw. einfach auf neue Geräte oder Anforderungen anpassen zu können. Zum anderen spielt die Erweiterung der Möglichkeiten selbst (Verfolgung weiterer Objekte, weitere Demontagevorgänge) eine tragende Rolle. Dies ist im Hinblick auf die notwendigen, komplexen und verschiedenartigen Regelungen einer in Zukunft geplanten vollständigen Demontage eines Motorraumes unumgänglich.

4.4.1 Flexible, universelle Einsetzbarkeit

In [Tonko 97] lag der Schwerpunkt auf der Entwicklung und Implementation eines Regelungsmechanismus unter Verwendung der zur Verfügung stehenden Geräte. Während dieser Entwicklung war dadurch eine möglichst geräteunabhängige Programmierung zweitrangig, wodurch automatisch starke Verflechtungen zwischen Programmkern und Geräteansteuerungen entstehen mußten. Dadurch ergibt sich die Situation, daß seither das Programm ausschließlich auf dem eingesetzten Steuerrechner bei eingeschalteten MiniVISTAs bzw. Robotern läuft. Da aus Zeitgründen stets Halbbilder der MiniVISTAs verwendet wurden, ist zudem sowohl der Bildeinzug als auch die Verfolgung daran speziell angepaßt.

Um nun bei einem eventuellen Ausfall von Geräten (Steuerrechner, Werkzeug- oder Beobachtungsroboter, Transputer, MiniVISTAs, Schnittstellenkarten) einen reibungslosen Fortgang der Forschungen soweit wie möglich zu gewährleisten sowie das System flexibel auch für reine Verfolgungsaufgaben verschiedener Art einsetzen zu können (z.B. zum risikolosen Test der Verfolgung), wird das Programmsystem *Visdom* so modifiziert, daß es unabhängig von den Geräten allein mit Bildfolgen von Platte auch auf anderen Rechnern arbeiten kann – und solche Bildfolgen auch aufzeichnen kann. Dazu wird das System insbesondere auf andere, nicht speziell für die Geräteansteuerungen konfigurierte Rechner (Sun Sparc Ultra II, Sun Sparc Ultra I, Sparc Station 10/20, Pentium PCs) portiert. Zudem wird das System übersetzbar gemacht für die im Rahmen der hier aufgeführten Untersuchungen relevanten Übersetzer (*gcc 2.7/2.8.1/2.95.2, SC 4.2*).

Der Bildeinzug wird reimplementiert, damit auch Vollbilder eingezogen werden können, und *Visdom* entsprechend für die Verwendung von Vollbildern angepaßt bzw. auf die Verwendung beliebiger Bildformate erweitert.

Diese Maßnahmen erlauben nun neben dem risikolosen Test des Verfolgungsmechanismus auch die Verfolgung spezieller Geräte in einem durchgeführten Projekt. Sie bilden aber auch die Voraussetzung der in Abschnitt 4.5.1 durchgeführten Optimierung hinsichtlich des am besten geeigneten Übersetzers und machen die in Abschnitt

4.5.2 folgende verteilte Berechnung der Lokalisation zur Beschleunigung erst möglich. Die Verbesserung hinsichtlich verarbeitbarer Bilder bildet zudem die Grundlage der in Abschnitt 4.1.7 verwendeten Vollbilder des Schiffsteils sowie der Verarbeitung der dort verwendeten hochauflösenden Bilder (vgl. Seite 106).

4.4.2 Regelung eines Schraubers

Bei einer Erweiterung von *Visdom* um einen weiteren Regelungsmechanismus bietet es sich an, die in [Gengenbach 94] vorgestellte Regelung eines Schraubers mit den in *Visdom* bestehenden Ansätzen zu Lokalisation, Verfolgung und Regelung zu realisieren. Diese Regelung wird aus mehreren Gründen ausgewählt. Zum einen stehen die in [Gengenbach 94] verwendeten Geräte und Objekte (Schrauberwerkzeug, Roboter, aufzuschraubendes Objekt) original zur Verfügung und wurden bereits in der zitierten Arbeit als regelungstaugliche Komposition nachgewiesen. Zum zweiten bildet die Aufschraubaktion einen prinzipiell anderen Fall einer Regelung als die bisher untersuchte Aktion zum Greifen eines Objektes. Zudem spielen Schrauben natürlich im Motorraum eines Fahrzeuges bei der Demontage eine wichtige Rolle. Das aus zwei Metallquadern unterschiedlicher Größe bestehende, von zwei Schrauben zusammengehaltene Objekt – in dieser Arbeit kurz ‘Doppelklotz’ genannt, vgl. Abbildung 4.17 – bildet daher optimale Grundlagen zur Erforschung der beim Schrauben auftretenden Effekte. Diese Regelung eignet sich zudem für die weiteren Forschungen besonders gut, weil diese im Gegensatz zur Greiferregelung monokular funktioniert und eine hinter dem Schrauber montierte Kamera auf der Roboterhand verwendet wird (Auge-auf-Hand-Konfiguration). Diese gravierenden Unterschiede zwischen den beiden Regelungen erbringen (durch die dadurch erzielte Verschiedenartigkeit der Prozesse, vgl. Abschnitt 4.2) den Nachweis eines hohen Grades an Regelungsrobustheit, da bei beiden Regelungsprozessen weitgehend die gleichen Routinen verwendet werden.

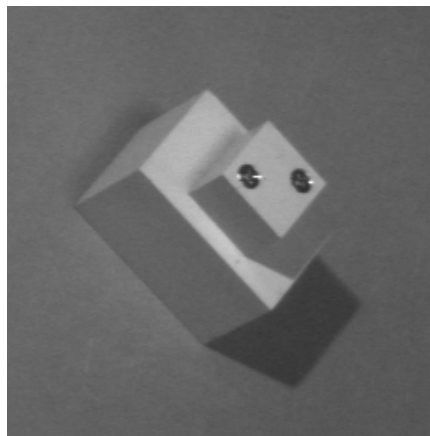


Abbildung 4.17: Doppelklotz aus [Gengenbach 94].

Zur Initialisierung der Schrauberregelung wird die Lokalisation aus Abschnitt 4.1 eingesetzt (anstatt der bei [Gengenbach 94] verwendeten, trinokularen und merkmalsbasierten Lageschätzung des Objekts [Gengenbach et al. 96]; eine Zusammenfassung findet sich auf Seite 13).

Während bei der Regelung des Greifers zur Batterieentnahme Sollwerte für die Relativlage zwischen Objekt und Werkzeug vorgegeben werden, müssen bei der Führung des Schraubers Sollwerte für die Relativlage der Kamera bezüglich des Doppelklotzes vorgegeben werden, da hier nur ein Objekt verfolgt wird. Zu minimieren ist folglich die Differenz zwischen der Istposition der Kamera bezüglich des Objekts und der Sollposition.

Dadurch daß bei der Schrauberregelung die Kamera bewegt wird, ergeben sich sehr große Änderungen der Verhältnisse im Bild. Um eine robuste Regelung des Schraubers realisieren zu können, wird die Werkzeugroboterposition während der Regelung abgefragt und zur Prädiktion der Position des Doppelklotzes⁸ eingesetzt (bei der Greiferregelung wird dagegen keine Bewegungsprädiktion durchgeführt).

Die Regelung des Schraubers geschieht in mehreren Teilschritten. Zuerst erfolgt eine grobe Anfahrt des Schraubers an den Doppelklotz aus über einem Meter Entfernung. Eine Feinregelung zur genauen Positionierung des Schraubers über der ersten Schraube schließt an. Nach deren Abschluß wird der Schrauber eingeschaltet, in den Schraubenschlitz eingetaucht und mit geeigneter Geschwindigkeit wieder zurückbewegt. Analog wird anschließend die zweite Schraube behandelt. Diese Abfolge zeigt, daß die Regelung in mehreren Sequenzen durchzuführen ist. Bisher wurde bei der Regelung des Greifers zur Batterieentnahme jedoch nur eine Sequenz benötigt. Zur Erhöhung der Flexibilität wird daher die Regelung bei dieser Gelegenheit so erweitert, daß bei durchzuführenden Regelungen bequem solche Regelungssequenzen vorgegeben werden können. Dies findet natürlich zum einen hier Anwendung. Bei der Regelung des Greifers dient die Erweiterung dazu, nach erfolgter Entnahme der Batterie diese neben dem Versuchsfahrzeug abzulegen – schließlich sollen die Demonteverfahren in Zukunft so ausgebaut werden, daß *mehrere* Objekte dem Motorraum entnommen und zweckgerecht abgelegt werden können.

Zur Illustration zeigt Abbildung 4.18 Bilder aus der Sicht der eingesetzten Kamera mit den eingeblendeten Lageschätzungen des Doppelklotzes. Der Schrauber ist am unteren Bildrand, etwas rechts von der Mitte als dunkler Fleck zu erkennen. Die Verfolgung funktioniert auch bei einer Variation der Beleuchtungsbedingungen 7 Pixel-robust und das Lösen der beiden Schrauben ist robust bzgl. der auf Seite 114 spezifizierten Anforderungen. Die Regelung funktioniert sogar so robust, daß der Doppelklotz während des Verfahrens des Schraubers manuell bewegt werden kann.

⁸Da sich die Kamera bewegt und die Position des Doppelklotzes bezüglich des Kamerakoordinatensystems geschätzt wird, ergibt sich eine zu schätzende Relativbewegung des Doppelklotzes bezüglich der Kamera.

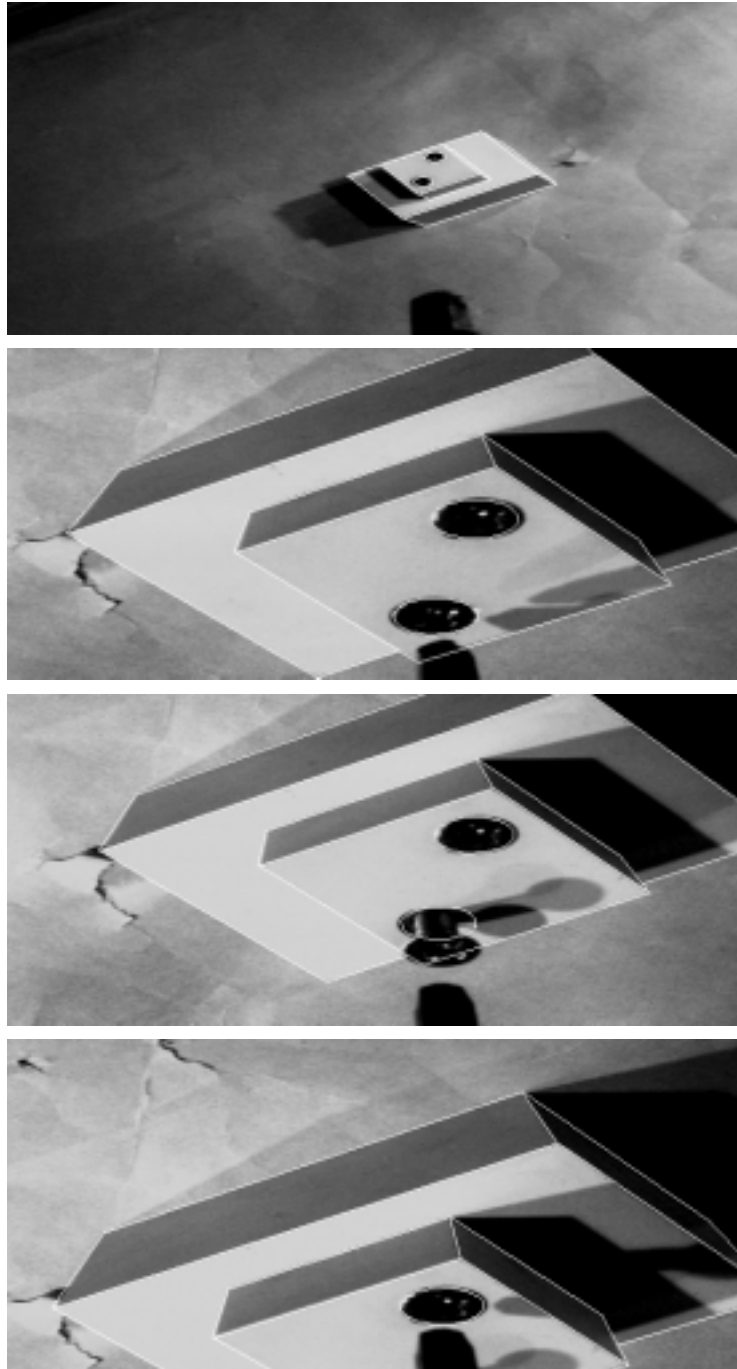


Abbildung 4.18: Lageschätzung des Doppelklotzes bei der Regelung des Schraubers zu den Halbbildzeitpunkten 1 (oberes Bild), 810 (zweites Bild von oben; vor dem Einschalten und Absenken des Schraubers), 1180 (drittes Bild von oben; nach dem Lösen der ersten Schraube) und 1439 (unteres Bild; vor dem Einschalten und Absenken des Schraubers). Zum Erreichen der im zweiten Bild dargestellten Zielposition mußte sich der Endeffektor, auf dem Kamera und Schrauber unbeweglich montiert sind, entsprechend translatorisch und rotatorisch bewegen.

4.4.3 Verfolgung weiterer Objekte

Bei den bisherigen Entwicklungen stand im Vordergrund, ein robustes Verfolgungsverfahren zu realisieren, vgl. [Tonko 97] sowie Abschnitt 4.2. Nachdem für die hierbei betrachteten Objekte zufriedenstellende Ergebnisse erzielt worden sind, ist im Hinblick auf die in Zukunft geplanten Demontageaktionen an vielen Motorraumteilen wichtig, den Verfolgungsprozeß an weiteren Objekten auszutesten. Dadurch können evtl. vorhandene Schwächen der Verfolgung aufgefunden bzw. beseitigt werden und aufgrund des vielfältigeren Testmaterials ein höherer Grad an Robustheit nachgewiesen werden. Dazu eignen sich insbesondere Objekte besonders gut, die den Verfolgungsprozeß vor neue Herausforderungen stellen. Das diesbezüglich geeignetste Objekt ist ein kugelförmiger Unterdruckbehälter für die Servolenkung im hinteren Teil des Motorraumes, vgl. Abbildung 4.19 (siehe auch die linke untere Ecke von Abbildung 1.2, links neben dem Luftfilter). Das Teil ist kugelförmig und besitzt um die Äquatorlinie einen Wulst. Im Gegensatz zu den bisher verwendeten Objekten ergibt sich daraus ein bei der Schätzung nicht eindeutig bestimmbarer Freiheitsgrad: die Drehung um die Polachse. Einziger Teil am Objekt, der eine Bestimmung dieses Freiheitsgrades zuläßt, ist eine rechteckige Lasche am oberen Pol. Aufgrund des möglicherweise bei der Verfolgung nur schwer bestimmbareren Freiheitsgrades sowie der rotationssymmetrischen, kreisförmigen Silhouette des Objektes im Bild ist mit Verfolgungsschwierigkeiten zu rechnen.

Das zweite geeignete Objekt findet sich im Rahmen eines Projektes, bei dem u.a. als Anforderung gilt, ein von einer Schiffswerft geliefertes Schiffsteil während einer Kreisbewegung der Kamera zu verfolgen. Das Schiffsteil weist mehrere Merkmale in Form von unterschiedlich geformten Löchern und von integrierten Schienen auf, wie sie bei den in der Werft verschweißten Teilen auftreten. Durch die Dichte der Details im Schiffsteil und die Verschiedenartigkeit der Merkmale stellen sich andere Anforderungen an den Verfolgungsprozeß als bisher, was das Teil für die Untersuchungen interessant macht.

Entgegen der Erwartungen kann sowohl die Verfolgung des kugelförmigen Behälters als auch des Schiffsteils im mono- wie auch binokularen Fall ohne Verfahrensmodifikationen robust ausgeführt werden. Abbildung 4.20 illustriert die monokulare Anpassung des kugelförmigen Behälters und Abbildung 4.21 die Stereoverfolgungen der Teile. Die monokulare Verfolgung des Schiffsteils während einer kreisförmigen Bahn der Beobachtungskamera zeigt Abbildung 4.24 im nächsten Abschnitt.

Durch die hiermit erfolgte Erweiterung der Verschiedenartigkeit des Testmaterials ist die eingesetzte Verfolgung als noch robuster nachgewiesen als bisher. Die oben erwähnten Anforderungen an die Verfolgung stellen keine neuen Schwierigkeiten dar.



Abbildung 4.19: Blick in den Motorraum des Versuchsfahrzeugs. Der kugelförmige Unterdruckbehälter für die Servolenkung ist im Bild in der linken, unteren Ecke zu sehen. In der rechten, oberen Ecke des Bildes kann man die Batterie erkennen und etwa in der Mitte des Bildes befindet sich der Luftfilter.

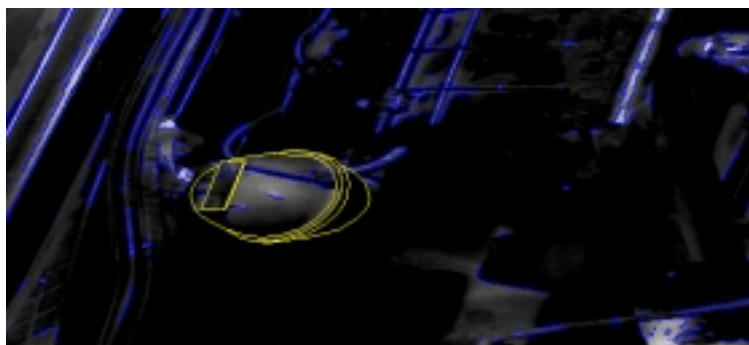


Abbildung 4.20: Monokulare Verfolgung des kugelförmigen Behälters.

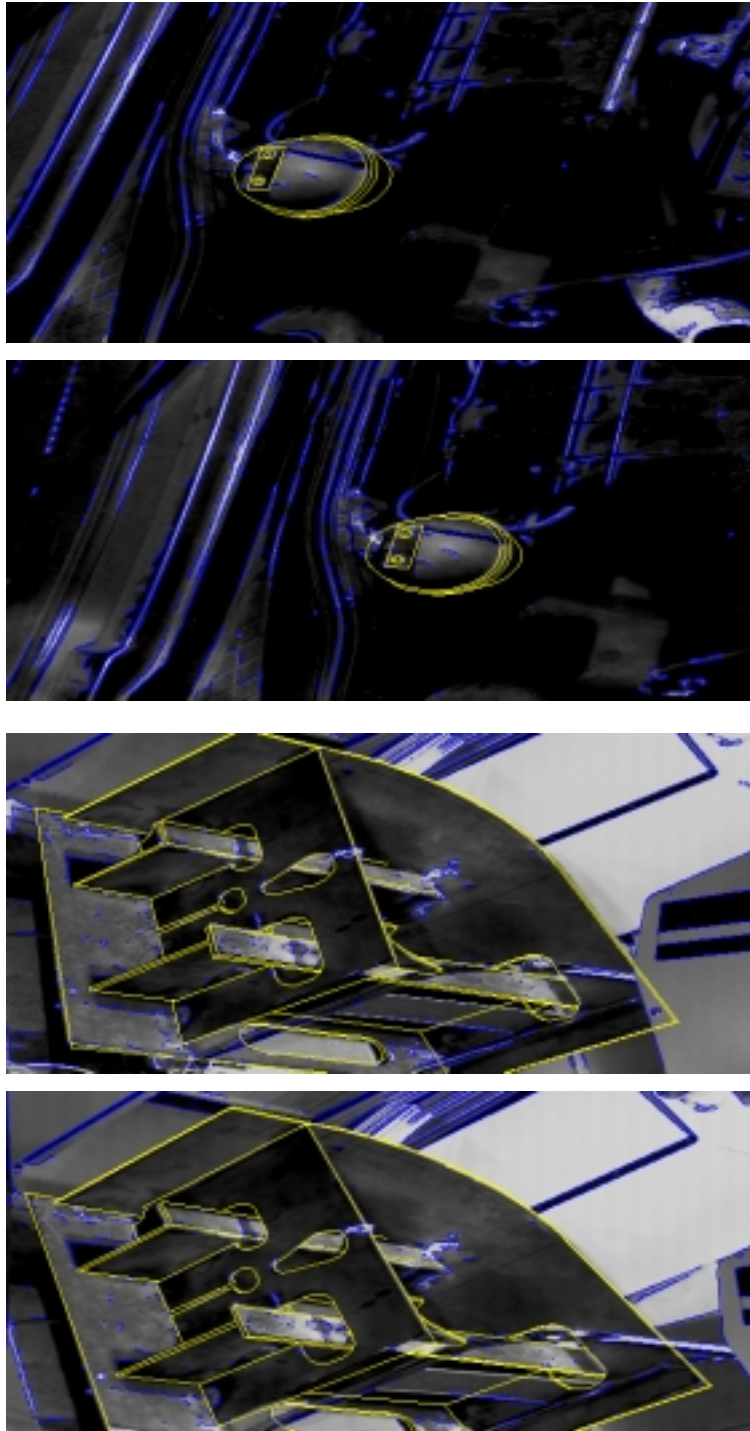


Abbildung 4.21: Stereoverfolgung des kugelförmigen Behälters (obere zwei Bilder, linke und rechte Kamera) und des Schiffsteils (untere zwei Bilder, oben linke und unten rechte Kamera).

4.4.4 Ansteuerung frei wählbarer räumlicher Bahnkurven des Beobachtungsroboters

Bisher hat es bei der Werkstückverfolgung ausgereicht, den Beobachtungsroboter auf Geradensegmenten über dem Motorraum periodisch vorwärts und rückwärts fahren zu lassen. Dies wurde insbesondere dazu genutzt, spezielle, für den Verfolgungsprozeß kritische Geradensegmentbahnen zu wählen, um die Schwachstellen des Verfolgungsprozesses gezielt analysieren und verbessern zu können, vgl. [Tonko 97]. Nachdem der Verfolgungsprozeß ein hinreichendes Stadium an Robustheit erreicht hat, tritt nun ein flexibles Einsatzspektrum des Verfolgungssystems in den Vordergrund. Damit verbunden ist die Anforderung, den Beobachtungsroboter auf beliebigen Bahnen über dem Motorraum fahren lassen zu können (etwa, um in Zukunft *während der Regelung* jeweils optimale Beobachtungspositionen anzufahren. Durch die Erhöhung der Verschiedenartigkeit des Testmaterials wird zugleich die Robustheit der Verfolgung besser nachgewiesen, vgl. Abschnitt 3.1.1). Bei dem Ganzen ist zu beachten, daß das gewünschte Objekt während der Fahrt stets im Bild der Kamera sichtbar sein muß und ‘glatte’ Bewegungen der Kameras ausgeführt werden, da ruckartige Bewegungen oder Richtungsänderungen der Kamera oft zu sehr großen Änderungen im Bild führen, denen der Verfolgungsprozeß nicht gewachsen ist.

Zur Lösung der Aufgabe soll *Visdom* so erweitert werden, daß allgemeine Splinekurven im Raum abgefahren werden können. Die Bahnkurven müssen dabei für einen sinnvollen praktischen Einsatz folgende Kriterien erfüllen:

- Die Kurven sollten C^2 -stetig sein, um ruckfreie Bewegungen bzw. Richtungsänderungen der Kameras zu gewährleisten.
- Die Festlegung einer solchen Kurve sollte möglichst einfach geschehen. Hier bietet es sich an, Stützpunkte auf der Kurve vorzugeben, da diese einfach durch Anfahren der entsprechenden Roboterpositionen bestimmbar sind.
- Damit das zu beobachtende Objekt während der gesamten Fahrt im Sichtfeld der Kamera bleibt, muß die Kurve entsprechend konstruiert sein.
- Durch den im vorhergehenden Abschnitt erwähnten Anwendungsfall muß es möglich sein, mit den Splinekurven zumindest näherungsweise einen Kreis zu modellieren.

Aufgrund dieser Kriterien bietet es sich an, gemäß Abbildung 4.22 vorzugehen (vgl. [Farin 94]). Die Punkte \bullet , \circ und \square (und damit alle Kurvenpunkte) repräsentieren dabei Roboterpositionen aus \mathbb{R}^6 (die ersten drei Parameter einer Roboterposition repräsentieren die Position, die letzten drei Parameter die Orientierung des Roboter-Endeffektors), die direkt zur Ansteuerung des Beobachtungsroboters verwendet werden. Die Bézier-Punkte \bullet werden als Stützpunkte vorgegeben und die Bézier-Punkte \circ

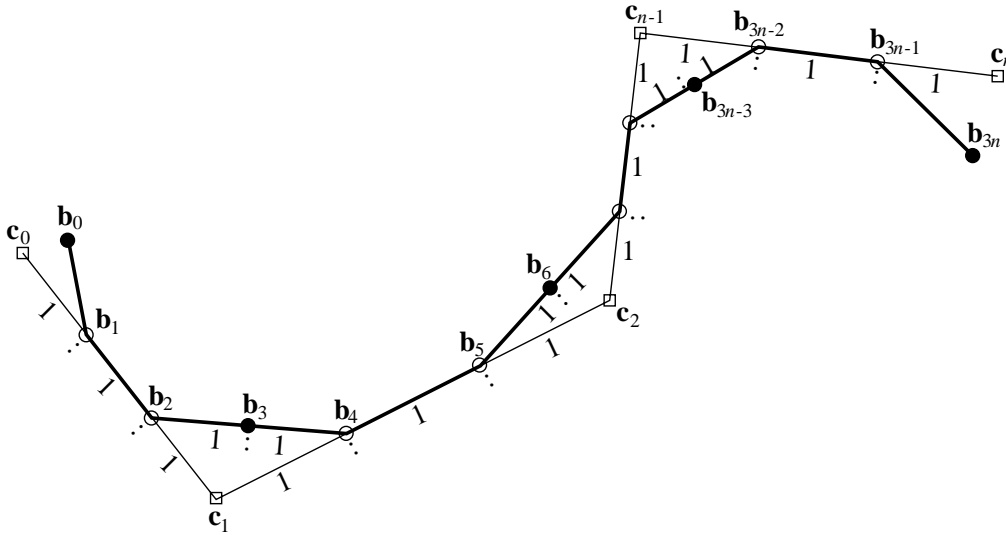


Abbildung 4.22: Splinekurve für die Fahrt des Beobachtungsroboters. Die Bézier-Punkte \bullet definieren die Stützpunkte der abzufahrenden Kurve und repräsentieren – wie die anderen Punkte auch – Roboterpositionen aus \mathbb{R}^6 . Die Bézier-Punkte \circ und die Kontrollpunkte \square des zugehörigen C^2 -stetigen uniformen Spline können daraus berechnet werden. Die einzelnen Punktketten $\bullet\text{---}\circ\text{---}\circ\text{---}\bullet$ bilden jeweils Bézier-Polygone vom Grad 3.

gemäß Abbildung daraus berechnet (die mathematischen Zusammenhänge folgen unten). Die einzelnen Punktketten $\bullet\text{---}\circ\text{---}\circ\text{---}\bullet$ bilden jeweils Bézier-Polygone vom Grad 3. Nach Konstruktion entsteht eine C^2 -stetige, uniforme Splinekurve mit den Kontrollpunkten \square (vgl. [Farin 94]).

Zwischen den Bézier-Punkten \mathbf{b}_i und den Kontrollpunkten \mathbf{c}_j bestehen gemäß Abbildung 4.22 folgende Beziehungen:

$$\mathbf{b}_{3i-1} = \frac{1}{3}\mathbf{c}_{i-1} + \frac{2}{3}\mathbf{c}_i, \quad (4.14)$$

$$\mathbf{b}_{3i+1} = \frac{2}{3}\mathbf{c}_i + \frac{1}{3}\mathbf{c}_{i+1}, \quad (4.15)$$

$$\mathbf{b}_{3i} = \frac{1}{2}(\mathbf{b}_{3i-1} + \mathbf{b}_{3i+1}) = \frac{1}{6}\mathbf{c}_{i-1} + \frac{2}{3}\mathbf{c}_i + \frac{1}{6}\mathbf{c}_{i+1},$$

für jeweils zulässige Indizes $3i - 1 \in \{0, \dots, 3n\}$, $3i + 1 \in \{0, \dots, 3n\}$ bzw. $3i \in \{0, \dots, 3n\}$. Aus der letzten Gleichung folgt

$$\begin{pmatrix} \mathbf{b}_3 \\ \mathbf{b}_6 \\ \vdots \\ \mathbf{b}_{3n-3} \end{pmatrix} = \begin{pmatrix} \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & 0 & \cdots & 0 \\ 0 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & \ddots & \vdots \\ \vdots & \ddots & & \ddots & & 0 \\ 0 & \cdots & 0 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{pmatrix} \begin{pmatrix} \mathbf{c}_0 \\ \mathbf{c}_1 \\ \vdots \\ \mathbf{c}_n \end{pmatrix} .$$

Die Verwendung von Vektoren innerhalb der Vektorklammern soll dabei andeuten, daß die Beziehung für jede Koordinate einzeln gilt.

- Ist eine *geschlossene Splinekurve* abzufahren, gilt $\mathbf{c}_0 = \mathbf{c}_n$ und $\mathbf{b}_0 = \mathbf{b}_{3n}$. Daraus folgt

$$\underbrace{\begin{pmatrix} \mathbf{b}_0 \\ \mathbf{b}_3 \\ \vdots \\ \mathbf{b}_{3n-3} \end{pmatrix}}_{=:\mathbf{b}} = \underbrace{\begin{pmatrix} \frac{2}{3} & \frac{1}{6} & 0 & \cdots & 0 & \frac{1}{6} \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & 0 & \cdots & 0 \\ 0 & & \ddots & & \ddots & \vdots \\ \vdots & \ddots & & \ddots & & 0 \\ 0 & \cdots & 0 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ \frac{1}{6} & 0 & \cdots & 0 & \frac{1}{6} & \frac{2}{3} \end{pmatrix}}_{=:\mathbf{A} \text{ (} n \times n\text{-Matrix)}} \underbrace{\begin{pmatrix} \mathbf{c}_0 \\ \mathbf{c}_1 \\ \vdots \\ \mathbf{c}_{n-1} \end{pmatrix}}_{=:\mathbf{c}} . \quad (4.16)$$

- Für eine abwechselnd vorwärts und rückwärts abzufahrende *offene Splinekurve* wird von einem sog. natürlichen Spline ausgegangen mit Endkrümmung 0, d.h. $\mathbf{c}_0 = \mathbf{b}_0$ und $\mathbf{c}_n = \mathbf{b}_{3n}$:

$$\underbrace{\begin{pmatrix} \mathbf{b}_0 \\ \mathbf{b}_3 \\ \vdots \\ \mathbf{b}_{3n} \end{pmatrix}}_{=:\mathbf{b}} = \underbrace{\begin{pmatrix} 1 & 0 & \cdots & 0 \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & 0 & \cdots & 0 \\ 0 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & \ddots & \vdots \\ \vdots & \ddots & & \ddots & & 0 \\ 0 & \cdots & 0 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ 0 & \cdots & & 0 & 1 \end{pmatrix}}_{=:\mathbf{A} \text{ (} n+1 \times n+1\text{-Matrix)}} \underbrace{\begin{pmatrix} \mathbf{c}_0 \\ \mathbf{c}_1 \\ \vdots \\ \mathbf{c}_n \end{pmatrix}}_{=:\mathbf{c}} . \quad (4.17)$$

Beides führt auf ein lineares Gleichungssystem der Form $\mathbf{b} = \mathbf{A}\mathbf{c}$. Die Kontrollpunkte \mathbf{c} können folglich berechnet werden über die Beziehung

$$\mathbf{c} = \mathbf{A}^{-1}\mathbf{b} ,$$

womit sich nun die gesuchten Bézier-Punkte \circ nach den Gleichungen (4.14) und (4.15) berechnen lassen.

Damit ist die Kurve eindeutig festgelegt und die einzelnen Bézier-Kurven

$$\mathbf{p}_i(t) := \sum_{j=0}^3 \mathbf{b}_{3i+j} B_j^3(t), \quad i = 0, \dots, n-1$$

können abgelaufen werden (B_j^3 bezeichnet dabei das j -te Bernstein-Polynom vom Grad 3). Dazu werden die \mathbf{p}_i mittels des de Casteljau-Algorithmus so lange unterteilt, bis sowohl der Maximalabstand der Position als auch der der Orientierung zwischen zwei aufeinanderfolgenden Bézier-Punkten unter eine vorher festgelegte Schwelle fällt. So wird gewährleistet, daß keine zu großen Schritte bei der Bahnfahrt ausgeführt werden. Da die berechneten Bézier-Punkte unterschiedlich weit auseinander liegen, muß nun noch vermieden werden, daß zu dicht beieinanderliegende Bézier-Punkte zu einer zu langsamen Bewegung des Roboters, d.h. während der Fahrt zu einer ungleichmäßigen Bewegungsgeschwindigkeit führen. Dazu wird die erhaltene Liste von Bézier-Punkten abgelaufen und zu jedem dieser Punkte der jeweils am weitesten entfernte genommen, dessen Abstand noch unterhalb der oben genannten Schwelle liegt. Dies definiert eine Liste von Punkten, die nacheinander anzufahren sind, wobei an den Endpunkten der Kurve abhängig davon, ob eine offene oder geschlossene Kurve abzufahren ist, unterschiedlich vorgegangen werden muß. Um den Rechner während der Verfolgung nicht unnötig zu belasten, wird diese Punktliste bei der Initialisierung des Beobachtungsroboters vorberechnet.

Beide Kurvenarten (4.16), (4.17) wurden implementiert und erfolgreich getestet. Bei allen getesteten Bahnkurven werden jeweils alle eingangs gestellten Forderungen erfüllt. Damit ist der Nachweis der Funktionstüchtigkeit des Verfahrens erbracht. Da die Verfolgung jeweils einwandfrei funktioniert, ist gleichzeitig die Verfolgung als entsprechend robust nachgewiesen (durch Vergrößerung der Testdatenmenge bzw. durch die erzielte Verschiedenartigkeit des Bildmaterials entsprechend robuster als bisher, vgl. S. 40).

Abschließend zeigt Abbildung 4.23, wie die auf Seite 133 benötigte Kreisbewegung der Kamera bei der Verfolgung des Schiffsteils mittels Vorgabe von drei symmetrisch angeordneten Bézier-Punkten \bullet realisiert werden kann. Die Festlegung der Kreisbahn zur Schiffsteilverfolgung geschieht, indem der Roboter in der Zelle an drei möglichst symmetrisch liegende Raumpunkte gefahren wird, bei denen das Schiffsteil ganz von der rechten Kamera erfaßt wird, und die drei abgelesenen Roboterpositionen als Stützpunkte \bullet der Kurve vorgegeben werden. Abbildung 4.24 zeigt zur Illustration die Anpassung des Schiffsteils bei der Kreisbewegung der Kamera zum ersten Halbbild (Lokalisationsergebnis), nach einem Drittel sowie nach zwei Dritteln der Fahrt. Nach Absolvierung des Vollkreises ergibt sich eine ähnliche Anpassung des Schiffsteils wie im oberen Teil der Abbildung. Zum Nachweis der Robustheit der Verfolgung wurde eine signifikante partielle Verdeckung des Schiffsteils vorgesehen (sichtbar im linken Drittel des unteren Teilbildes von Abbildung 4.24). Die Verfolgung bleibt während der gesamten Fahrt so

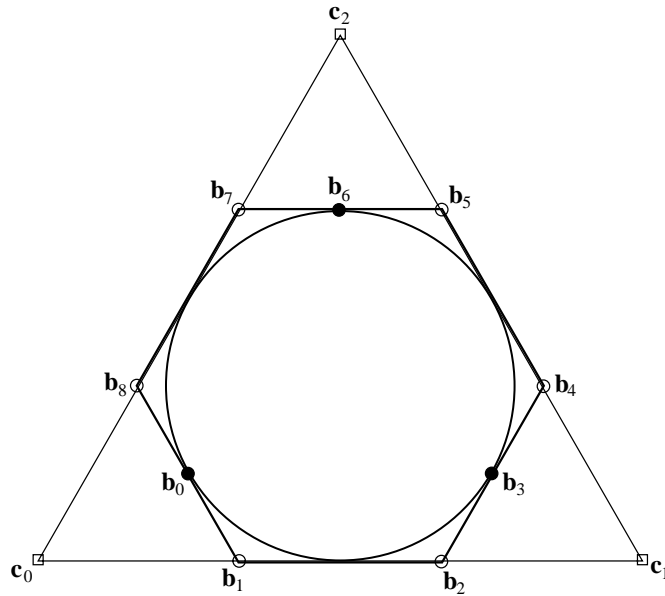


Abbildung 4.23: Spezielle Splinekurve für die Kreisbewegung des Beobachtungsroboters mit Stütz- und Kontrollpunkten. Drei vorgegebene, symmetrisch angeordnete Bézier-Punkte \bullet definieren eine Kreisbahn. Wie in Abbildung 4.22 bezeichnen die Punkte \square die Kontrollpunkte des C^2 -Spline und die Punktketten $\bullet\text{---}\circ\text{---}\circ\text{---}\bullet$ legen die einzelnen Bézier-Kurven vom Grad 3 fest.

genau wie zu den dargestellten drei Zeitpunkten, was die Robustheit des Verfahrens gegenüber partiellen Verdeckungen und der gewählten Kamerabahn belegt.

4.5 Geschwindigkeit

Dem Geschwindigkeitsaspekt in Form von Echtzeitanforderungen kommt bei der sichtsystemgestützten automatischen Demontage während des eigentlichen Regelungsprozesses ein besonders hoher Stellenwert zu, schließlich muß die Bildauswertung mit den sich kontinuierlich bewegenden Robotern schritthalten können. Darüber hinaus müssen Ausnahmefälle zum Schutz der Geräte so schnell wie möglich erkannt werden. In natürlicher Weise ergibt sich als Anforderung an den Regelungsprozeß, möglichst jedes von den Kameras gelieferte Bild schritthaltend mit dessen Eintreffen im Rechner auszuwerten.

Da bei der Lokalisation zur Initialisierung des zugrundeliegenden Verfolgungsprozesses keine Roboter bewegt werden, spielt der oben genannte Aspekt hierbei keine Rolle. Aus diesem Grund wurde ja die Auswahl des Verfahrens zur Lokalisation ausschließlich nach Robustheitskriterien getroffen, vgl. Abschnitt 4.1. Da die Robustheit des Verfahrens nachgewiesen werden konnte, tritt an dieser Stelle der Praktikabilitätsaspekt aus der Sicht des Anwenders in den Vordergrund: da die Lokalisation signifikante Rechenzeiten

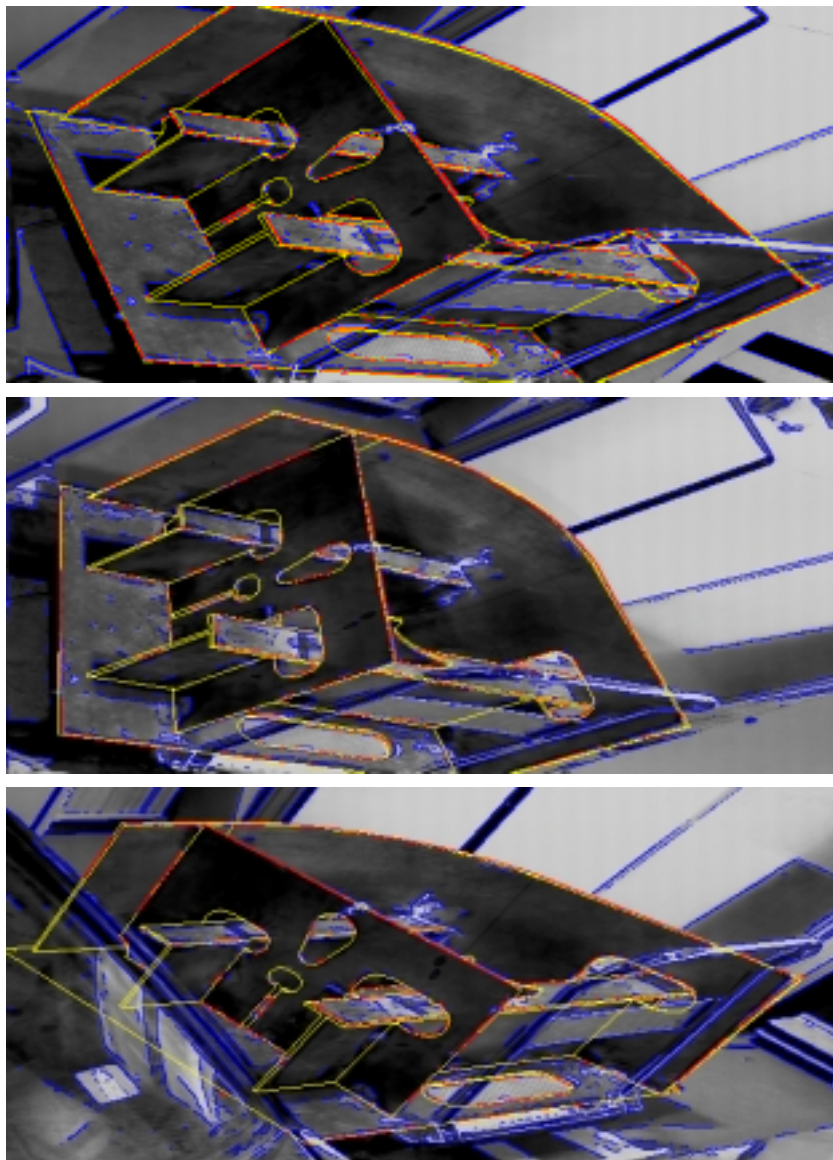


Abbildung 4.24: Ergebnis der Anpassung des automatisch lokalisierten Schiffsteils bei einer Kreisbewegung der Kamera zu den Halbbildzeitpunkten 1 (oben; Beginn der Kreisbahn), 387 (Mitte; Kreis zu einem Drittel absolviert) und 774 (unten; Kreis zu zwei Dritteln absolviert). Nach Vollendung der Kreisfahrt ergibt sich eine ähnliche Anpassung wie die im oberen Teil dargestellte zum Halbbildzeitpunkt 1.

zwischen 3 und 20 Minuten je Objekt benötigt, ist es im Zuge der mit der Zeit anfallenden zahlreichen Experimente wünschenswert, die Rechenzeiten zu minimieren und damit die Weiterentwicklung der Regelung zu beschleunigen.

4.5.1 Steigerung der Echtzeitfähigkeit bei der Regelung

Einleitend wurde der Stellenwert von Echtzeitanforderungen an das Regelungssystem beschrieben. Bevor diesbezüglich Verbesserungen durchgeführt werden können, muß als Erstes spezifiziert werden, nach welchen Kriterien dies geschehen soll.

Die *Zykluszeit* bezeichne die Zeit, die zwischen der Beendigung der Auswertung eines Halbbildes und der des nächsten verstreicht. Die *Latenzzeit* sei definiert als die Zeit zwischen dem Zeitpunkt, bei dem ein von Kamera oder Festplatte eingelesenes Halbbild im Speicher des Rechners zur Verfügung steht, bis zu dem Zeitpunkt, bei dem die Auswertung dieses Halbbildes abgeschlossen ist. Die Halbbilder inklusive Kantenelementbilder werden von einer MiniVISTA (vgl. S. 6) mit einer Zykluszeit etwas unter 40 *ms* geliefert. Die gewünschte Echtzeitanforderung, jedes gelieferte Bild auswerten zu können, kann daher abgebildet werden auf die Aufgabe, die Zykluszeit des Regelungsmechanismus zu senken auf unter 40 *ms*. Zur Erfüllung allein dieser Anforderung wäre eine Fließbandstruktur denkbar, welche durch eine geschickte Parallelisierung die gegebene Zeitschranke für die Zykluszeit zu unterschreiten vermag, aber gleichzeitig hohe Latenzzeiten nach sich zieht. Da es bei der Regelung jedoch wichtig ist, mittels geeigneter Stellsignale für die Roboter schnell auf Gegebenheiten in der Szene reagieren zu können, soll als weitere zu erfüllende Anforderung zugleich eine weitestgehende Minimierung der Reaktionszeit des Systems, d.h. der Latenzzeit gelten.

Da der binokulare Verfolgungsprozeß komplexer ist als der monokulare, bietet es sich als ersten Schritt an, die *monokulare* Verfolgung in obigem Sinn echtzeitfähiger zu gestalten. Im Folgenden wird zur Bewertung der durchgeführten Maßnahmen die monokulare Verfolgung der Batterie verwendet, da sie unter allen monokularen Verfolgungen die zeitaufwendigste ist.

Das Regelungssystem nach [Tonko 97] arbeitete auf der GIGAmachine mit einer Zykluszeit von etwa 160 *ms* und einer Latenzzeit um 240 *ms* (zur Gerätekonfiguration siehe Abbildung 5.3 auf Seite 107 in [Tonko 97]). Es muß also eine Beschleunigung von mindestens einem Faktor von 4 realisiert werden. Um einen so hohen Faktor erreichen zu können sind mehrere Maßnahmen erforderlich, wie das Folgende zeigt. Eine Optimierung der zentralen Verfolgungsroutinen wurde bereits im Rahmen der Arbeiten zu [Tonko 97] durchgeführt und scheidet daher im Weiteren aus.

Von einer Umstellung auf einen schnelleren Rechner kann allem voran die deutlichste Verringerung beider Zeiten (Zyklus- und Latenzzeit) erwartet werden. Das System nach [Tonko 97] lief auf einer Rechnerkonfiguration von insgesamt 150 *MHz*, vgl. [Tonko 97]. Von einer Portierung auf eine zu diesem Zweck angeschaffte SUN Sparc Ultra II mit zwei 300 *MHz*-Prozessoren (zum Kaufzeitpunkt schnellster angebotener Rechner) ist – durch Vergleich der Taktfrequenzen – mit einer Beschleunigung von etwa dem gewünschten Faktor zu rechnen. Nach Durchführung der Portierung ergibt sich die in Abbildung 4.29 auf Seite 149 dargestellte Gerätekonfiguration (vgl. Neuerschaltung der Spezialrechner nach Abschnitt 4.3). Unter Verwendung des auf dem neuen Rechner installierten Übersetzers `gcc 2.8.1` (unter Einsatz der höchsten Optimierungstufe

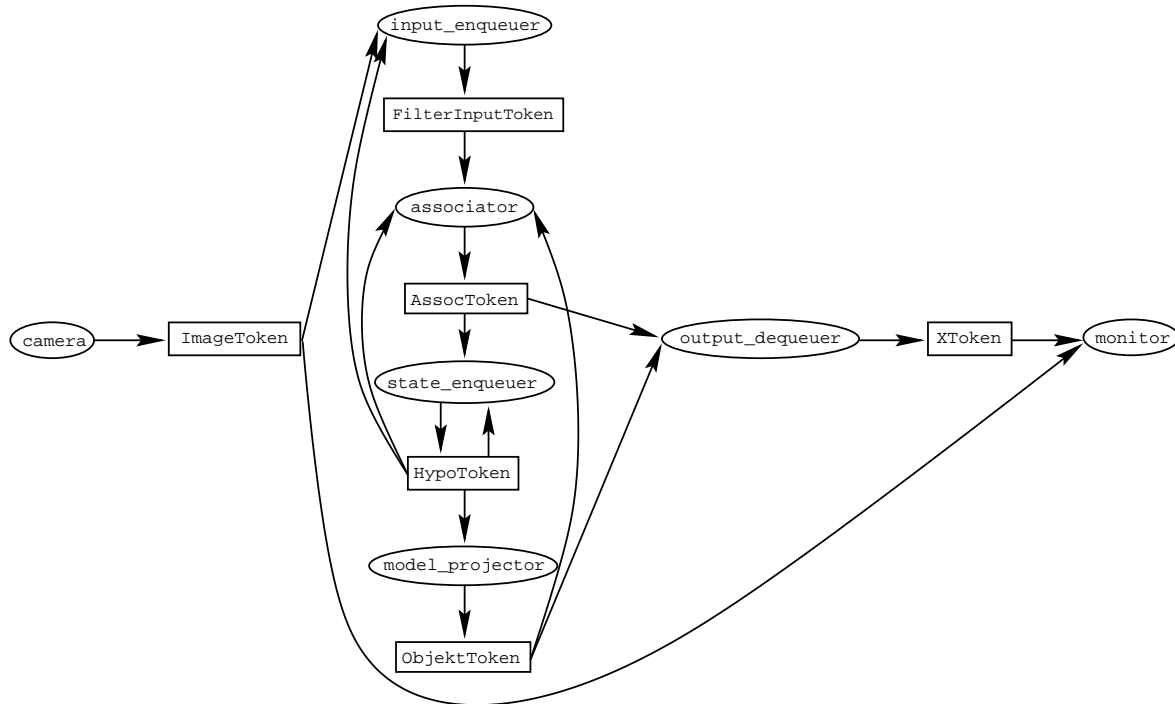


Abbildung 4.25: Datenfluß- und Ablaufdiagramm der prozeß(faden)orientierten monokularen Verfolgung nach [Tonko 97]. Datenelemente in den Warteschlangen (rechteckig gezeichnet): `ImageToken` enthält das von der Kamera gelieferte Bild inkl. Kantenelementbild, `AssocToken` die im Anpassungsverfahren benutzten Kantenelemente, `HypoToken` u.a. den Zustandsvektor des Anpassungsverfahrens und die Kovarianzmatrix, `ObjektToken` die projizierten Modellkanten und `XToken` die Informationen zur Darstellung auf dem Bildschirm. Bedeutung der Prozeßfäden (mit Ellipsen dargestellt): `camera` realisiert den Bildeinzug, `input_enqueueuer` berechnet eine Lauflängenkodierung der Kantenelementbilder, `associator` ordnet die Kantenelemente den projizierten Modellkanten zu, welche von `model_projector` berechnet werden, `state_enqueueuer` realisiert den erweiterten Kalman-Filter, `monitor` stellt die Ergebnisse auf dem Bildschirm dar und `output_dequeueuer` bereitet diese vorher geeignet auf.

beim Übersetzen) verkürzt sich die Zykluszeit um einen Faktor von 3,88. Die erreichte Zykluszeit liegt mit $41,2\text{ ms}$ noch knapp über der zu unterbietenden Bildeinzugszeit von 40 ms , vgl. Zusammenfassung in Tabelle 4.3. Die Latenzzeit liegt deutlich über der Zykluszeit. Dies liegt an der in [Tonko 97] eingeführten Ablaufstruktur des Verfolgungsprozesses, die *prozess(faden)-orientiert* arbeitet. Abbildung 4.25 verdeutlicht dieses Ablaufschema: Zur Realisierung einer möglichst effizienten Ausführung auf der GIGAmachine wurde die Verfolgung im Rahmen der Arbeiten zu [Tonko 97] in mehreren, parallel laufenden Prozeßfäden realisiert und deren Datenaustausch über Warteschlangen geeignet geregelt.

Aufgrund der Warteschlangenverwaltung ist zu erwarten, daß dieses Konzept auf dem

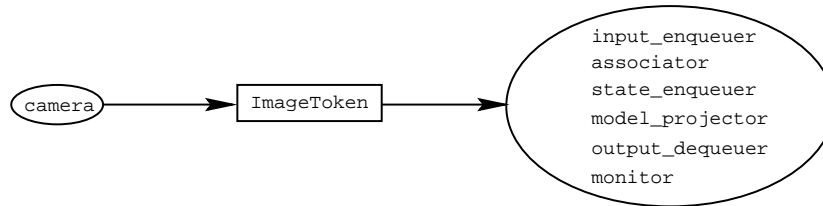


Abbildung 4.26: Realisierte Prozedurstruktur bei der monokularen Verfolgung.

	Zykluszeit	Latenzzeit	Verbesserung	
			Zykluszeit	Latenzzeit
Ursprüngliches System n. [Tonko 97]	160 <i>ms</i>	240 <i>ms</i>	–	–
Portierung auf Ultra II (gcc 2.8.1, Prozeßstruktur)	41,2 <i>ms</i>	59,5 <i>ms</i>	74,3%	75,2%
Realisierung als Prozedurstruktur	42,7 <i>ms</i>	37,4 <i>ms</i>	-3,6%	37,1%
Ablauf in der Echtzeitklasse	41,7 <i>ms</i>	34,5 <i>ms</i>	2,3%	7,8%
Bereinigung des Programmsystems	38,6 <i>ms</i>	34,6 <i>ms</i>	7,4%	-0,3%
Optimierung mittels eingesetztem Übersetzer (SC 4.2)	38,6 <i>ms</i>	29,0 <i>ms</i>	0,1%	16,2%

Tabelle 4.3: Schritte bei der Verbesserung der Echtzeitanforderungen an die monokulare Verfolgung, gemessene Zyklus- und Latenzzeiten sowie die bei jedem Schritt erzielte Verbesserung bezogen auf den vorhergehenden Schritt. Erläuterungen siehe Text. Zur Prozentangabe bei der Zykluszeit ist anzumerken, daß der Wert anhand der genauen Zykluszeiten ermittelt worden ist, und nicht mit den angegebenen Rundungswerten.

verwendeten Rechner zu unnötigen Verzögerungen führt. Eine Durchführung in einem hauptsächlich *prozedural* organisiertem Ablauf verspricht daher auf dem verwendeten Rechner eine kleinere Zyklus- bzw. Latenzzeit. Diese Erwartung entspringt der Tatsache, daß ein Prozeßfaden nur zu diskreten Zeitpunkten vom Betriebssystem gestartet wird. Außerdem müssen die Prozeßfäden auf die beiden zur Verfügung stehenden Prozessoren verteilt werden, wodurch Umschaltzeiten anfallen. Zur Prüfung dieser Hypothese wird die in Abbildung 4.25 gezeigte Prozeßstruktur in eine geeignete Prozedurstruktur transformiert, bei denen die bisherigen Prozeßfäden als Prozeduren ausgelegt und anstatt des Warteschlangenkonzepts Übergabeparameter für die Prozeduren verwendet werden. Lediglich eine Warteschlange zur Parallelisierung von Bildeinzug und Verfolgung wird noch verwendet, siehe Abbildung 4.26. Die in Tabelle 4.3 dargestellten Ergebnisse belegen die Verbesserung der Latenzzeit um 37,1% auf 37,4 *ms*. Sie liegt damit deutlich unterhalb der Bildeinzugszeit, was eindeutig für die Verwendung der Prozedurstruktur spricht. Allerdings verschlechtert sich die Zykluszeit – wenn auch unwesentlich – gegenüber der Verwendung der Prozeßstruktur.

Um die Zykluszeit zu verringern, wird der Zeitaufwand für die Verwaltung der noch be-

stehenden Warteschlange reduziert, indem die sogenannte Echtzeitklasse (eine Möglichkeit, höchste Echtzeitanforderungen von Seiten des Betriebssystems zuzusichern) des verwendeten Betriebssystems *Solaris 2.6* verwendet wird. Dies vermag zwar die genannte Verschlechterung der Zykluszeit zumindest zum Teil wieder auszugleichen und die Latenzzeit um 7,8% zu verringern, jedoch ist das Ziel einer Zykluszeit unterhalb von 40 *ms* noch nicht erreicht. Um dieses Ziel zu erreichen, erweist sich die Bereinigung und Verbesserung des Warteschlangenkonzeptes sowie des Bildeinzugs samt der Reimplementation einiger Teile als geeignet. Unter ‘Bereinigung’ ist hierbei das Entfernen nicht mehr benötigter Teile gemeint, vgl. Abschnitt 4.3. Diese Teile wurden im Rahmen der Untersuchungen von [Tonko 97] für diverse Versuche mit dem Bildeinzug auf der GIGAmachine benötigt, als es darum ging, überhaupt ein laufendes System zu entwickeln. Tabelle 4.3 belegt, daß dadurch die Zykluszeit mit 34,6 *ms* endlich – wie angestrebt – unter 40 *ms* zu liegen kommt. Messungen ergeben, daß dies der genauen Bildeinzugszeit durch das Kamera/MiniVISTA-System entspricht. Da die Verfolgung aufgrund der unterhalb dieser Schwelle liegenden Latenzzeit in jedem Zyklus auf das Bildmaterial von der Kamera warten muß, ist die Zykluszeit der Modellanpassung gleich der Zykluszeit des Bildeinzugs (vgl. Abb. 4.26).

Zur Erfüllung der eingangs definierten Anforderung, die Latenzzeit so weit wie möglich zu minimieren, verbleiben zwei Möglichkeiten innerhalb des Programms, die an dieser Stelle untersucht werden. Zum einen handelt es sich um die Durchführung einer Lauflängenkodierung der Kantenelementbilder vor der Zuordnung der Kantenelemente zu den projizierten Modellkanten (siehe Bildunterschrift zu Abbildung 4.25). Es stellt sich die Frage, ob dies nur auf der GIGAmachine eine Beschleunigung der Berechnungen gebracht hat. Zum zweiten wird der Algorithmus zur Zuordnung von Kantenelementen auf mehrere Prozeßfäden verteilt und parallel ausgeführt. Eingestellter Wert ist hier die Anzahl von 4 verwendeten Prozeßfäden. Weder die Abschaltung der Bildkompression (sie ermöglicht eine effizientere Berechnung der Kantenelementzuordnungen; die Latenzzeit verschlechtert sich bei Abschaltung um 15,8%) noch eine Änderung der Anzahl an eingesetzten Prozeßfäden (bereits optimal eingestellt) liefert geringere Latenzzeiten.

Eine signifikante Verringerung der Latenzzeit ist nach Betrachtung obiger Möglichkeiten zum gegenwärtigen Stand nur noch von einer Optimierung hinsichtlich des am besten geeigneten Übersetzers zu erwarten. Daher wird der diesbezüglich im besten Ruf stehende Übersetzer *SC 4.2* der Firma *SUN* verwendet ([Leuck 00]). Wie Tabelle 4.3 zu entnehmen ist, läßt sich die Latenzzeit alleine durch die Verwendung dieses Übersetzers um weitere 16,2% verringern.

Es ist anzumerken, daß sich die obigen Ausführungen im Wesentlichen auch auf den binokularen Fall übertragen. Allerdings gestaltet sich hier die Überführung der Prozeßfadenstruktur in eine Prozeßstruktur etwas komplizierter, da insgesamt drei Prozesse (nicht Prozeßfäden) verwendet werden (eingeführt im Rahmen von [Tonko 97]): je einer pro Kamera und einer, der die Ergebnisse der beiden miteinander verrechnet, vgl.

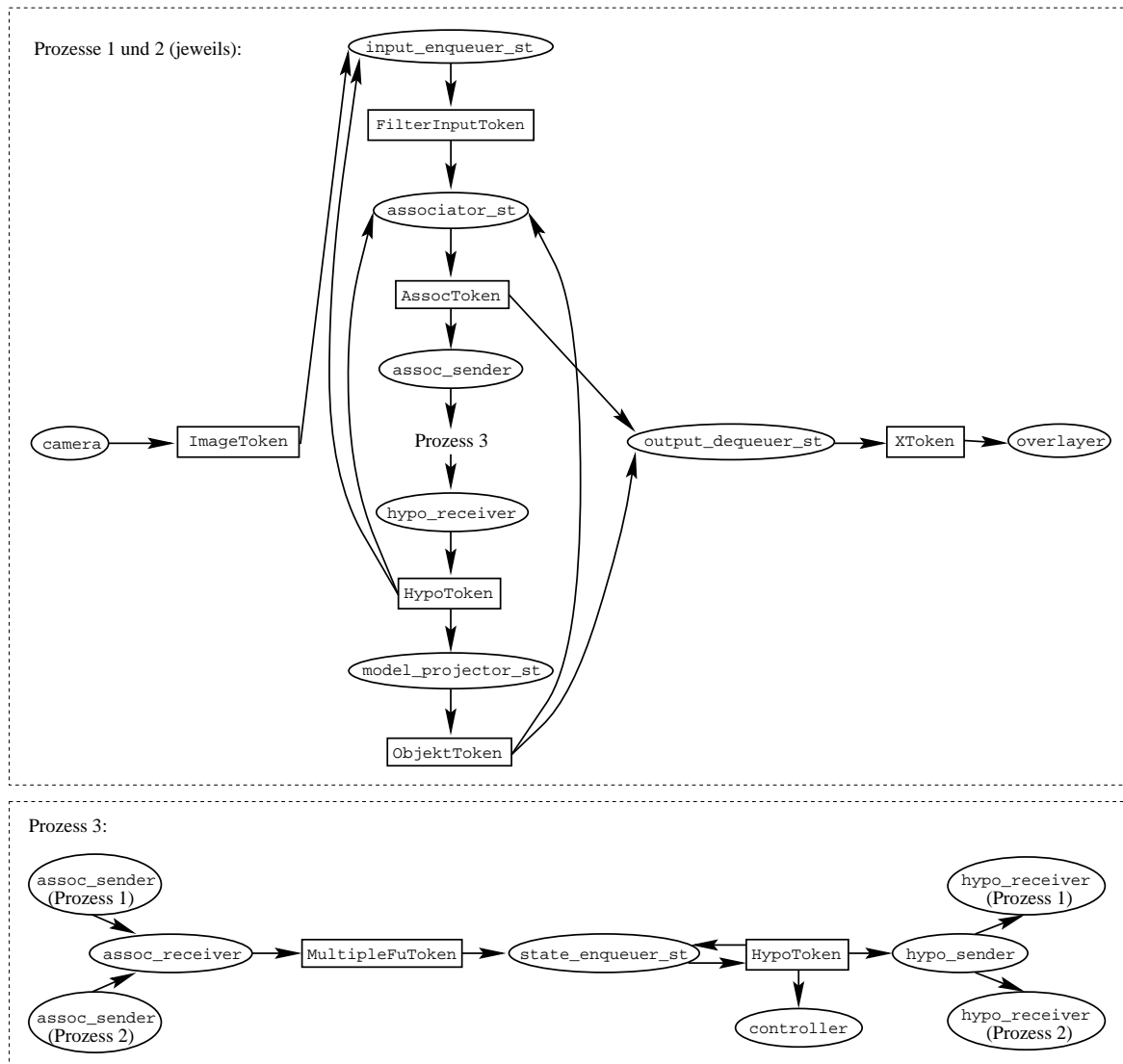


Abbildung 4.27: Zusammenspiel und Ablauf der drei Prozesse bei der Stereoanpassung nach [Tonko 97]. Zwei Prozesse realisieren jeweils den im oberen Teil dargestellten Ablauf, der dritte Prozeß den unten dargestellten. Die Darstellungen folgen dem monokularen Fall aus Abb. 4.25. Im Gegensatz zum monokularen Fall kommen Prozeßfäden zur Interprozeßkommunikation `assoc_sender/receiver` bzw. `hypo_sender/receiver` hinzu. `controller` bezeichnet die Roboterregelung, die in Abb. 4.25 nicht eingezeichnet wurde. Im Gegensatz zur monokularen Verfolgung übernimmt hier aus historischen Gründen `input_enqueueuer_st` auch die Darstellung der Bilder auf dem Bildschirm.

Abbildungen 4.27 und 4.28. Die Verwendung von Prozeßfäden anstatt der drei Prozesse bringt keine Verbesserung der Ausführungszeiten, obwohl Umschaltvorgänge zwischen Prozeßfäden allgemein als deutlich kürzer gelten als Umschaltzeiten zwischen Prozessen. Von einer eventuellen Prozeduralisierung ist ähnliches zu erwarten. Als Unterschied

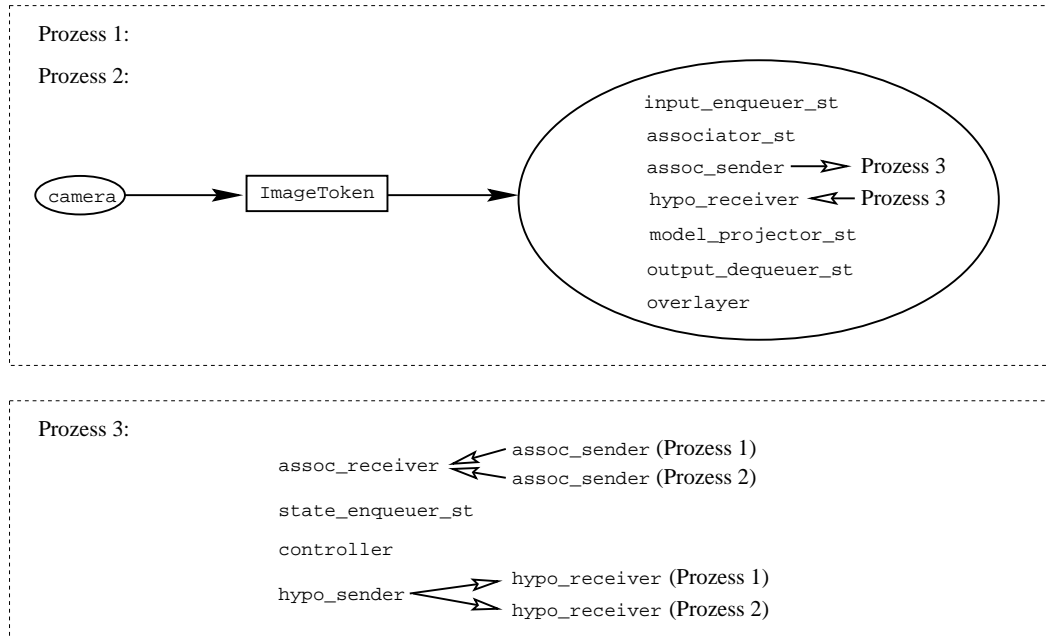


Abbildung 4.28: Realisierte Prozedurstruktur bei der binokularen Verfolgung.

zum monokularen Fall ist ferner festzuhalten, daß der Übergang auf den Übersetzer **SC 4.2** im Vergleich zur monokularen Verfolgung einen höheren Gewinn bei der Zykluszeit (3,3%), aber zugleich einen niedrigeren Gewinn bei der Latenzzeit (6,7%) bringt.

Abschließend zeigt Tabelle 4.4 die Zyklus- und Latenzzeiten sowie die jeweilige Verteilung der Ausführungszeiten auf die einzelnen Programmteile für alle monokularen und binokularen Verfolgungen bzw. Demontagevorgänge. Alle monokularen Vorgänge haben eine Zykluszeit von maximal 38,5 *ms*, d.h. *jedes* von der Kamera eingezogene Bild wird schritthaltend ausgewertet (zum Vergleich: bei [Tonko 97] war es nur jedes dritte). Im deutlich aufwendigeren binokularen Fall wird immerhin noch mehr als jedes zweite Bild der Kameras verwendet. Dadurch kann nun die Regelung deutlich schneller als bisher auf äußere Gegebenheiten reagieren, um Risiken für die eingesetzten Geräte zu vermeiden, die Demontage schneller als bisher ablaufen zu lassen und die Regelung deutlich robuster als bisher zu gestalten.

Zusammenfassend kann festgehalten werden, daß Beschleunigungen im Zusammenhang mit der Erfüllung von Echtzeitanforderungen oft nur durch die Betrachtung mehrerer Ebenen erzielt werden können. Im hier dargestellten konkreten Fall hat sich neben der Verwendung eines schnelleren Rechners oder der Optimierung auf Programm- oder Parameterebene vor allem die Optimierung hinsichtlich des am besten geeigneten Übersetzers bewährt. Wie sich gezeigt hat, kann allein dadurch die Ausführungszeit um bis zu 16% beschleunigt werden. Die Untersuchungen belegen außerdem, daß der Durchbruch zur Unterschreitung der kritischen Schwelle von 40 *ms* bei der Zykluszeit

	Zykl.	Latenz.	cam	inp	assc	state	modp	outp	mon
Batterie (mono)	38,4	26,6	38,4	3,9	11,2	0,2	10,7	0,7	2,4
Luftfilter (mono)	38,3	14,6	38,4	4,0	5,0	0,2	5,0	0,4	2,7
kugelf. Beh. (mono)	38,4	16,4	38,4	3,9	4,3	0,2	7,7	0,3	2,5
Schiffsteil (mono)	38,4	17,2	38,4	4,0	8,4	0,2	4,1	0,5	2,9
Batterie (stereo)	60,2	52,6	39,8	24,5	13,9	3,9	9,8	0,5	27,3
Luftfilter (stereo)	46,0	34,9	41,5	9,0	15,7	5,2	4,3	0,7	11,9
kugelf. Beh. (stereo)	48,3	39,0	41,4	10,9	11,0	4,8	12,0	0,4	14,4
Schrauberregelung	38,5	14,0	38,3	3,8	3,3	4,2	2,5	0,3	2,8
Greiferregelung	71,4	71,5	43,6	31,2	16,5	12,4	10,7	0,8	26,3

Tabelle 4.4: Mittlere Zyklus- und Latenzzeiten sowie durchschnittliche Dauern der einzelnen Programmteile in *ms* (die Abkürzungen stehen für *camera*, *input_enqueuer*, *associator*, *state_enqueuer+controller*, *modelprojector*, *output_dequeuer* und *monitor/overlayer*, vgl. Abb. 4.26 und 4.28) für die monokularen (mono) und binokularen Verfolgungen (stereo), die (monokulare) Schrauberregelung zur Doppelklotzdemontage und die (binokulare) Greiferregelung zur Entnahme der Batterie aus dem Motorraum. Bezüglich der hohen Zeiten in der letzten Spalte ist zu bemerken, daß sich hier Wartezeiten auf den Bildeinzug widerspiegeln.

letztendlich nur durch die Bereinigung des historisch gewachsenen Programmsystems möglich wurde. Dieser Punkt sei an dieser Stelle besonders betont, um herauszuheben, welche Auswirkungen sich aufgrund von Programmsystemen ergeben, die mit der Zeit – als ganz natürlicher Teil der Entwicklung – unterschiedlichen Anforderungen und Randbedingungen genügen müssen (vgl. besonders Abschnitt 4.3).

4.5.2 Verteilte Berechnung der Lokalisation

Da bereits bei der Entwicklung der Lokalisation relativ hohe Rechenzeiten abzusehen waren, wurden zeitkritische Teile bereits bei der Implementation effizient realisiert und programmtechnische Optimierungen durchgeführt. Zur Beschleunigung der Lokalisation ist daher auf dieser Ebene kaum noch etwas zu gewinnen. Da das Verfahren aus Robustheitsgründen nicht modifiziert werden soll, muß mehr Rechenkapazität zur Verfügung gestellt werden, um die durchzuführenden Operationen in kürzerer Zeit bewerkstelligen zu können. Eine Portierung auf einen mittlerweile erhältlichen, im Vergleich zum bisher verwendeten etwa vier mal so schnellen Rechner läßt zwar eine Reduzierung der Lokalisationszeiten um etwa 75% erwarten, jedoch steht eine Möglichkeit in Aussicht, wie eine weitaus stärkere Verringerung der Zeiten erzielt werden kann. Wie das geschehen kann, wird schnell klar, wenn man sich sämtliche am Institut verwendbare Geräte vergegenwärtigt (Abbildung 4.29). Bereits zur Beschleunigung der Verfolgung

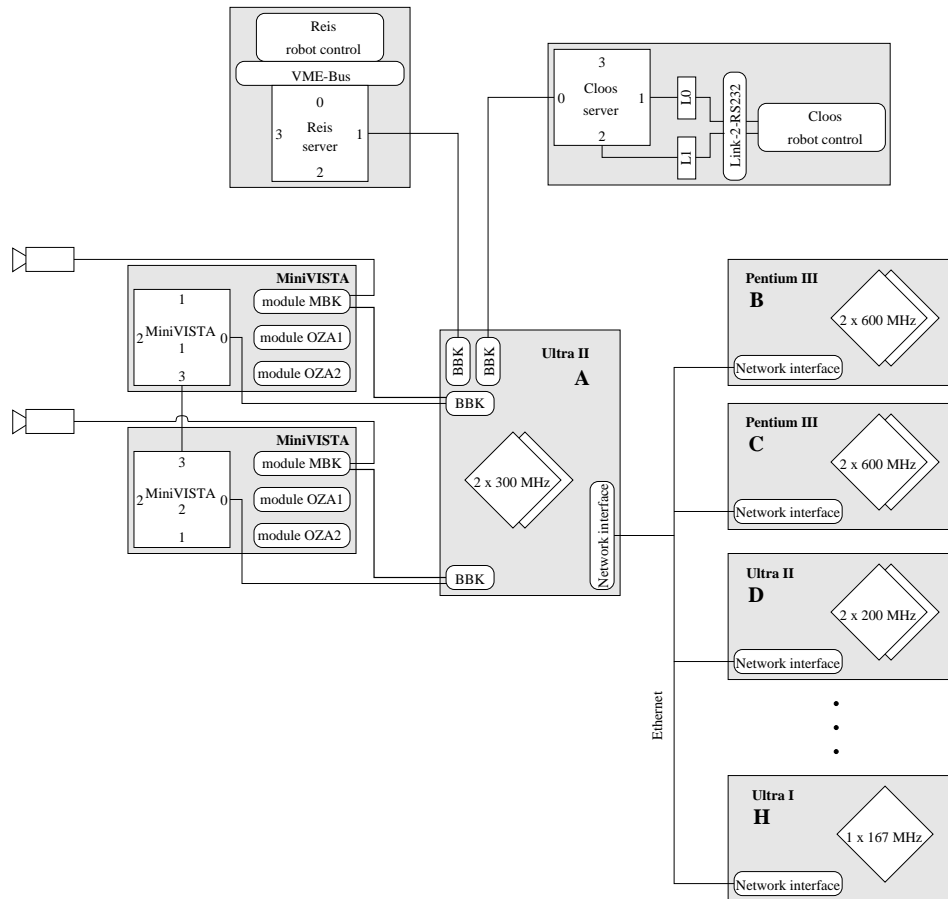


Abbildung 4.29: Sämtliche relevante Geräte des Instituts. In der Mitte ist der Steuerrechner dargestellt. Man erkennt auf der linken Seite die beiden MiniVISTAs für Bildeinzug und Kantenelementberechnung. Der oben links dargestellte Spezialrechner dient der Ansteuerung des Beobachtungsroboters, der oben rechts der des werkzeugtragenden Roboters. Die im Laufe der Zeit zu unterschiedlichen Zwecken angeschafften Rechner des Instituts sind auf der rechten Seite abgebildet. Sie stellen hinreichend viel potentielle Rechenleistung zur Beschleunigung der Lokalisation zur Verfügung.

wurden bisher stets Zusatzrechner (MiniVISTAs) eingesetzt – eine Maßnahme, die es bisher überhaupt erst möglich machte, Echtzeitanforderungen in die Tat umsetzen zu können. Es bietet sich an, zur Beschleunigung der Lokalisation in ähnlicher Weise nicht nur den Steuerrechner zu verwenden. Die Lokalisation soll daher zur Beschleunigung auf allen Rechnern verteilt ausgeführt werden. Grundlage dafür bietet das alle Rechner verbindende 10 MBit Ethernet sowie das allen Rechnern gemeinsame Betriebssystem Solaris, das alle dafür benötigten Werkzeuge zur Verfügung stellt. Da Visdom auf allen Rechnern lauffähig ist – die Grundlage dafür wurde in Abschnitt 4.4.1 geschaffen – und das Betriebssystem von der Art des verwendeten Rechners abstrahiert, hat solch ein Vorhaben trotz der Inhomogenität des Rechnernetzes gute Aussichten auf Erfolg.

Rechner	Prozessortyp	Kapazität
<i>A</i>	Ultra II	2×300 MHz
<i>B, C</i>	Pentium III	2×600 MHz
<i>D, E</i>	Ultra II	2×200 MHz
<i>F, G, H</i>	Ultra I	1×167 MHz

Tabelle 4.5: Zur Verfügung stehende Rechner mit einer Kapazität über 150 MHz.

Aufgrund der Kapazität der drei schnellsten Rechner werden für die Verteilung alle Rechner mit einer Mindestkapazität über 150 MHz als praktikabel eingestuft⁹, siehe Tabelle 4.5. Es sei darauf hingewiesen, daß keiner der Rechner zu diesem Zweck angeschafft wurde. Alle Rechner wurden in verschiedenen historischen Situationen zu unterschiedlichen Zwecken angeschafft. Die beiden Pentiums wurden beispielsweise zuletzt gekauft, um in naher Zukunft die MiniVISTAs zu ersetzen.¹⁰ Sie liegen insbesondere bei der Lokalisation brach.

Da jede Suchstufe der Lokalisation aus z.T. vielen Lagebewertungen besteht, kann jeder Suchschritt in naheliegender, natürlicher Weise aufgeteilt werden. Nach der Bewertung der Lagehypothesen durch die zur Verfügung stehenden Prozessoren werden die einzelnen Teilergebnisse nach jeder Suchstufe auf dem Zentralrechner *A* gesammelt und ausgewertet. Anhand der Auswertung wird die nächste Suchstufe vorbereitet, die wiederum verteilt abläuft.

Die Verteilung selbst beinhaltet zwei Komponenten. Die eine Komponente betrifft die Verteilung der auf einem Rechner anfallenden Berechnungen auf die zur Verfügung stehenden Prozessoren innerhalb des Rechners. Die zweite Komponente umfaßt die Verteilung der bei der Lokalisation anfallenden Berechnungen auf die zur Verfügung stehenden Rechner. Ersteres wird umgesetzt durch Verteilung der Aufgabe eines Rechners auf zwei Prozeßfäden. Die zweite Komponente impliziert die Verwendung entsprechender Sockel-Kanäle zur Kommunikation zwischen dem Zentralrechner *A* und den Satellitenrechnern *B* bis *H*. Zur Initialisierung erhält jeder Rechner vom Zentralrechner das zugrundeliegende Kantenelementbild sowie die Parametrisierung der Lokalisation, um die einheitliche Datengrundlage bzw. Parametrisierung sicherzustellen. Die Parametrisierung der Verteilung wird so einfach wie möglich gehalten. Zur interaktiven Auswahl der bei der Verteilung zu berücksichtigenden Rechner sind einfach deren Namen in einer Liste anzugeben.

⁹Zum einen ist von Rechnern mit einer geringeren Kapazität kein lohnenswerter Betrag zur Beschleunigung zu erwarten und zum anderen stellen die zur Verfügung stehenden Rechner mit einer Kapazität unter 150 MHz nicht genügend Hauptspeicher zur Verfügung.

¹⁰Die Rechner haben eine Kapazität erreicht, bei der es erstmals möglich sein sollte, die in hochintegrierte Schaltkreise gegossene Kantenelementberechnung der MiniVISTAs programmtechnisch in nahezu der gleichen Geschwindigkeit durchführen zu können. Da die MiniVISTAs schon so alt sind, daß kaum Ersatzteile mehr zu beschaffen sind, ist ein rechtzeitiger Umstieg von großer Bedeutung.

Mit diesem Werkzeug kann nun der Einfluß der beteiligten Rechner auf die Lokalisationszeit studiert werden. Zum Vergleich der verschiedenen Rechnerkonfigurationen ist es dabei sinnvoll, ein Vergleichsmaß zur Verfügung zu haben. Zu diesem Zweck werden die einzelnen Taktfrequenzen der beteiligten Prozessoren zu einem Maß für die Anzahl an Operationen pro Sekunde zur Beschreibung der Kapazität der Gerätekonfiguration addiert. Dagegen mag es durchaus berechtigte Einwände geben. Ergebnisse in [Leuck 00] und [Nagel et al. 99] belegen jedoch bei der Verfolgung von Fahrzeugen in Straßenverkehrsszenen, daß die Ausführungsrate bei jedem Rechnertyp unter Verwendung eines festen Übersetzers einen linearen Zusammenhang zur Taktfrequenz aufweist. Zwar hat jeder Rechnertyp seine eigene Kennlinie, doch sei hier vereinfachend von näherungsweise gleichen Charakteristiken ausgegangen. Obiges Maß kann also durchaus als grobes Maß zum Vergleich der Leistung verschiedener Rechnerkonfigurationen dienen.

Abbildung 4.30 veranschaulicht die Abhängigkeit der Lokalisationszeit bzw. -rate¹¹ von der gewählten Rechnerkonfiguration. Dabei wurden beginnend bei dem Zentralrechner bzw. den schnellsten Rechnern nach und nach immer mehr Rechner hinzugenommen (Rechnerkonfigurationen: A (ein Prozeßfaden); A (zwei Prozeßfäden); A, B; A-C; A-D; A-E; A-H). Wie man erkennen kann, gehen die Zeiten ab einer Gerätekonfiguration von 3400 Millionen Operationen pro Sekunde in einen Sättigungsbereich über – was aufgrund der Kapazitäten der Rechner F bis H auch nicht verwunderlich ist. Außerhalb des Sättigungsbereichs zeigt die Lokalisationsrate ein näherungsweise lineares Verhalten, was als Rechtfertigung für oben angenommenes Leistungsmaß angesehen werden kann.

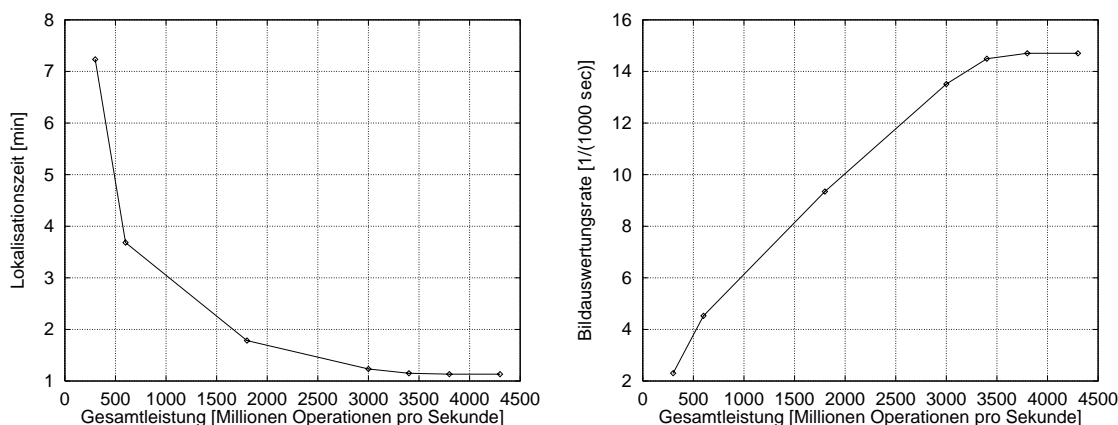


Abbildung 4.30: Abhängigkeit der Lokalisationszeit (links) bzw. -rate (rechts) von der Leistung der verwendeten Gerätekonfiguration bei einer Lokalisation des Schiffsteils. Das verwendete Schiffsteilbild wurde mit freundlicher Unterstützung von A. Bachem im neuen Technikum des Fraunhofer-Instituts für Informations- und Datenverarbeitung (IITB) von einer Hochdynamikkamera aufgenommen und zur Verfügung gestellt.

¹¹Gleich Kehrwert der Lokalisationszeit.

Objekt	A (1-fädig) 300 <i>MOpS</i>	A (2-fädig) 600 <i>MOpS</i>	A, B 1800 <i>MOpS</i>	A, B, C 3000 <i>MOpS</i>	A, B, C, D 3400 <i>MOpS</i>
Batterie	45:38	24:01	10:51	6:53	6:00
Luftfilter	6:47	3:43	1:37	1:01	0:55
kugelförmiger Behälter	6:17	3:14	1:30	0:57	0:52
Doppelklotz	2:51	1:23	0:43	0:29	0:27
Schiffsteil (Standardkamera)	9:15	4:53	2:26	1:32	1:25
Schiffsteil (hochauflösende Kamera)	20:04	12:02	6:23	4:11	3:59
Schiffsteil (Hochdynamikkamera)	7:14	3:41	1:47	1:14	1:09

Batterie	14:25	–	6:41	4:58	4:15
Luftfilter	9:52	–	3:49	2:31	2:11
kugelförmiger Behälter	10:35	–	4:29	2:44	2:30
Schiffsteil (Standardkamera)	10:10	–	4:55	3:04	2:48

Tabelle 4.6: Mono- (oberer Teil) und binokulare Lokalisationszeiten (unterer Teil) in *min:sek* abhängig von der Rechnerkonfiguration (Millionen Operationen pro Sekunde wurde abgekürzt mit *MOpS*). Anmerkung zur dritten Spalte, unten: Da die Stereolokalisation in Anlehnung an die Stereoverfolgung bereits in zwei Prozessen realisiert ist, machen Messungen bei Verwendung des Rechners *A* mit zwei Prozeßfäden keinen Sinn.

Tabelle 4.6 faßt die Ausführungszeiten aller mono- und binokularen Lokalisationen außerhalb des Sättigungsbereichs zusammen. Die maximal durch die Verteilung erzielbare Beschleunigung der Lokalisation (bei Konfiguration *A, B, C, D*) im Vergleich zur unverteilter Version bewegt sich zwischen einem Faktor von 3,4 bei der Stereolokalisation der Batterie und 7,6 bei der monokularen Lokalisation der Batterie. Das ist zwar wesentlich weniger als das Leistungsmaß erwarten läßt – gegenüber der einfädigen Lokalisation auf Rechner *A* hat diese Konfiguration bzgl. des gewählten Maßes etwa die 11,3-fache Leistung – aber die erhaltenen Lokalisationszeiten bewegen sich im Bereich zwischen 27 Sekunden und 6 Minuten, was wesentlich praktikabler ist als 3 bis 45 Minuten.

Die in diesem Abschnitt durchgeführten Untersuchungen belegen, daß es möglich ist, *in Situationen hohen Rechenaufkommens alle zur Verfügung stehenden Rechner – selbst in inhomogenen Rechnernetzen – zu akquirieren und zur Beschleunigung einer zeitaufwendigen Aufgabe zu nutzen*. Dies wurde im vorliegenden Fall an einer Lokisationsaufgabe demonstriert, um die Ausführungszeiten praktikabler zu machen.

4.6 Analysierbarkeit

Zur Verbesserung der Analysierbarkeit von auftretenden Fehlern in Programm und Gerätestruktur sei auf die Ausführungen in Abschnitt 4.3 verwiesen.

Bei der Analysierbarkeit der Ergebnisse können einige bei der Personenverfolgung bereits umgesetzte Konzepte zur Darstellung der Ergebnisse auf dem Bildschirm erfolgreich auf den Fall der sichtsistemgestützten Demontage übertragen werden. Zusätzlich konnte eine Verbesserung bezüglich einiger Ein- und Ausgabeformate erzielt werden. Im einzelnen wurden folgende Verbesserungen vorgenommen, deren Umsetzung sich zwar einfach gestaltete, deren Nutzbarkeit jedoch für die in dieser Arbeit dokumentierten Forschungen insbesondere beim Auffinden von Problemursachen wertvolle Dienste leistete.

- Von nun an flackerfreie Einblendung des projizierten Modells und der im Anpassungsprozeß verwendeten Kantenelemente,
- Einblendung zusätzlicher nützlicher Daten im Anzeigefenster, wie die Kantensuchbereiche und die gezählten Kantenelemente im Verlauf der Lokalisation, oder die farbliche Einblendung der im Verfolgungsprozeß nicht verwendeten Kantenelemente,
- Integration einer Lupe zum Vergrößern von dargestellten Ergebnissen ähnlich wie bei der Personenverfolgung (per Mausklick kann in einfacher Weise die Vergrößerungsstufe gewählt werden),
- Einstellung der Systemkovarianzen (vgl. Abschnitt 4.2.2) nicht mehr im Programm, sondern via Parameterdatei,
- Angabe von Ein- und Ausgabetransformationen nicht mehr in Form homogener Transformationsmatrizen (16 Parameter), sondern in Form äquivalenter Transformationsvektoren (6 Parameter; intuitiv wesentlich zugänglicher).

Die letzten beiden Punkte verbessern in natürlicher Weise insbesondere auch die *Bedienbarkeit* des Programms.

Dies zeigt deutlich, daß die Betrachtung des Analysieraspekts nicht nur Vorteile bei der Wartung des Programmsystems bringt, sondern automatisch als Teil der Entwicklung entsteht und die Verbesserung der anderen Aspekte befruchtet.

Kapitel 5

Gegenüberstellung der beiden Teilbereiche

Die sich ergebende Betrachtung des Automatisierungsaspektes bewirkte bei der sightssystemgestützten Demontage die Erweiterung um eine automatische Initialisierung des Verfolgungsprozesses. Dagegen war die Verfolgung von Personen bisher auf einem Stand, bei dem noch kein hinreichend geringer Benutzerinteraktionsgrad zu verzeichnen war. Daher bezog sich die Verbesserung der Automatisierung bei der Personenverfolgung noch auf die Verfolgung selbst. Eine Übertragung der Lokalisation von der Demontage auf den Bereich der Personenverfolgung wäre zudem aufgrund der zahlreichen zu bestimmenden Freiheitsgrade als nicht praktikabel einzustufen gewesen. Anstatt ein Lokalisationsverfahren aus der Literatur zu adaptieren oder ein für die Personenverfolgung geeignetes zu entwerfen, wurde es daher als wichtiger eingestuft, die Verfolgung selbst weiter zu verbessern, schließlich bildet die Objektverfolgung den Kern bei der Erkennung von Personenbewegungen. Da auf diesem Gebiet aufgrund der wesentlich komplexeren Anforderungen als bei der Demontage noch kein hinreichender Stand erreicht werden konnte, war es als wichtiger einzustufen, den Wesenskern zu verbessern, anstatt diesen zur Verbesserung der Automatisierung zu erweitern.

Die Betrachtung der Robustheit gestaltet sich in beiden untersuchten Bereichen als besonders vordergründig, da hier bei beiden die Hauptschwächen zu beobachten waren. Aufgrund der ähnlichen Funktionsweise bei beiden Verfahren erwies sich eine einheitliche Definition der Robustheit als sehr nützlich. In beiden Bereichen zeigte sich zudem, daß die auftretenden Robustheitsschwächen nicht monokausaler Natur sind, sondern vieler Teiluntersuchungen und -verbesserungen bedürfen. Dies läßt den Schluß zu, daß allgemein bei modellbasierten Sightssystemen – welche bereits einen gewissen Stand der Technik erreicht haben – in Sachen Robustheitsverbesserung grundsätzlich mit der Analyse und Modifikation zahlreicher Einzelfaktoren gerechnet werden muß.

Insbesondere zur Analyse von auftretenden Robustheitsschwächen erweisen sich geeignete Werkzeuge zu deren Untersuchung als sehr vorteilhaft. Da das Programm zur

Personenverfolgung in dieser Hinsicht bereits weit fortgeschritten war, konnte Einiges auf die Demontage übertragen werden.

Aufgrund des Einsatzes von Geräten ergab sich bei der Demontage im Gegensatz zur Personenverfolgung zusätzlich die Betrachtung von Aspekten wie Zuverlässigkeit, Sicherheit und Ausfallsicherheit. Durch den erreichten Stand bei der Verfolgung bzw. Lokalisation rückte außerdem bei der Demontage der Geschwindigkeitsaspekt immer weiter in den Vordergrund. Zudem sind bei der Roboterregelung Geschwindigkeitsanforderungen in Form von Echtzeitaspekten direkt mit Robustheits- und Sicherheitsaspekten gekoppelt. Da diese Punkte bei der Personenverfolgung sowie bei der Initialisierung des der Demontage zugrundeliegenden Verfolgungsprozesses nicht vorhanden sind, spielt dagegen dort die Ausführungsgeschwindigkeit eine untergeordnete Rolle. Daraus resultierte, daß auch komplexe Algorithmen zur Sicherstellung der Robustheit eingesetzt werden konnten (vgl. Abschnitte 3.1.3 und 4.1). Zunächst bietet sich hier lediglich aus Praktikabilitätsgründen eine Beschleunigung an (vgl. Abschnitt 4.5.2). Dagegen steht die Erfüllung der Echtzeitanforderungen bei der Demontage dem Einsatz zu komplexer Algorithmen zur Robustheitssteigerung entgegen, weshalb hier ein erweiterter Kalman-Filter und *kein* iterativ erweiterter Kalman-Filter mit genauerer Optimierung (siehe Personenverfolgung) zum Einsatz kommt. Zudem ist die Verfolgung bei der Demontage hinreichend robust, was den Einsatz des erweiterten Kalman-Filters rechtfertigt.

Sobald ein gewisser Robustheitsstandard erreicht ist, treten in natürlicher Weise Erweiterungsaspekte hinsichtlich Flexibilität und Anwendungsspektrum in den Vordergrund. Während bei der Demontage aus diesem Grund entsprechende Erweiterungen durchgeführt wurden, besteht bei der Personenverfolgung noch ein Robustheitsstand, bei dem andere Aspekte zum gegebenen Zeitpunkt wichtiger sind.

All diese Betrachtungen zeigen, wie sich die beiden Teilbereiche durch ihre Verschiedenartigkeit und die sich unterscheidenden Rahmenbedingungen ergänzen können. Dieser Punkt kann insbesondere auch anhand des folgenden Paradebeispiels zur Demonstration des Zusammenspiels der beiden Bereiche erläutert werden. Wie sich zeigen wird, kann dadurch ein interessanter Effekt modellbasierter Systeme erkannt werden.

In Abschnitt 3.1.2 wurde erläutert, daß sich aufgrund der rekursiven Wirkungsweise eines Verfolgungsprozesses Rechenungenauigkeiten bei der Personenverfolgung zu chaotischen Störungen der Anpassungsergebnisse aufschaukeln können. Zwar ergeben sich bei der Demontage keine Hinweise auf solch massive chaotische Auswirkungen, aber da der Verfolgungsprozeß ebenfalls rekursiv arbeitet, muß sich das Verhalten dynamischer Systeme auch hier auf irgendeine Art und Weise widerspiegeln. Der große Vorteil, der sich beim Transfer ausgewählter Untersuchungen von der Personenverfolgung auf die Demontage ergibt, liegt in den Möglichkeiten, welche durch die Versuchsumgebung der Roboterzelle bereitgestellt werden: Da durch den Beobachtungsroboter Kameratrajektorien vorgegeben und genau reproduziert werden können, kann untersucht werden, wie sich die Verfolgungsergebnisse in Abhängigkeit von einer immer wieder gleichen Rela-

tivbewegung zwischen Objekt und Kamera verhalten. Bei der Personenverfolgung wäre dies nur sehr schwer zu realisieren. Es ist zu erwarten, daß sich zumindest ansatzweise dynamisches Verhalten aus den Ergebnissen erkennen läßt.

Zur Durchführung der Untersuchungen werden drei bisher behandelte Objekte des Motorraumes (Batterie, Luftfilter und kugelförmiger Behälter) verfolgt, wobei die dabei verwendete Kamera eine lineare Bahn über dem Motorraum des Versuchsfahrzeugs beschreibt, welche periodisch zwei Mal abgefahren wird (eine Bewegung in Richtung des Objektes, auf dem gleichen Wege zurück, wieder auf das Objekt zu und wieder vom Objekt weg). Das bedeutet, daß die gleiche Bahn zwei Mal vorwärts und rückwärts abgefahren wird. Da die tatsächliche Relativposition des Objektes zur Kamera dabei den gleichen Verlauf nimmt, sollte die geschätzte Position $(x, y, z)^T$ des Objektes auch jeweils die gleichen Werte annehmen (bis auf zufallsverteilte Störungen durch Rauscheinflüsse im Bildmaterial).

Abbildung 5.1 zeigt die Positionsschätzungen $(x, y, z)^T$ der Objekte während der zwei Bewegungsperioden der Kamera. Zur übersichtlichen Darstellung der Ergebnisse wurden die Positionsschätzungen in der Abbildung jeweils in die xy - sowie die xz -Ebene projiziert. Bei Betrachtung der Grafiken fällt auf, daß die Trajektorien nicht immer aus nur einer Zusammenhangskomponente bestehen, sondern daß auch zwei Zusammenhangskomponenten auftreten können!¹ Dies tritt nicht nur bei einer der drei Verfolgungen auf, sondern bei allen, was als deutliches Indiz auf ein prinzipielles Phänomen zu werten ist. Das bedeutet, daß sich der Zustandsvektor bei jeder Periode der Kamerabewegung in verschiedenen Zustandsräumen bewegen kann, obwohl die Modellanpassung bei visueller Inspektion durch einen Betrachter gleich gut zu sein scheint. Offenbar kann die grobe Lage eines Objektes durch das aktuelle Modellanpassungsverfahren gut geschätzt werden, wohingegen bei der Bestimmung der genauen Lage ein Unschärfefaktor ins Spiel kommt. Die damit einhergehende maximale Schätzungsgenauigkeit kann aus den Abbildungen bestimmt werden zu weniger als 1 cm. Da sich der Abstand zwischen Objekt und Kamera im Bereich zwischen einem halben Meter und einem Meter bewegt, ist diese Schwankung im Rahmen des Toleranzbereichs.

Das bedeutet, daß selbst Effekte geringen Ausmaßes, wie Kamerarauschen und andere Rauscheinflüsse, im Hinblick auf den Zustandsraum, in dem sich die Lageschätzung eines Objekts im weiteren Verlauf einer Bildfolge bewegt, eine große Rolle spielen. Das bedeutet auch, daß dieser Effekt dem Verfolgungsverfahren inhärent ist, d.h. nicht vollständig beseitigt werden kann. Die eingangs erwartete Parallele zum Bereich dynamischer Systeme ist damit anhand der Verfolgung von Motorraumteilen belegt. Bei anderen modellbasierten Sichtsystemen ist dieser Effekt ebenfalls zu berücksichtigen.

¹Bei der Batterieverfolgung rühren die beiden Zusammenhangskomponenten von den beiden Bewegungsperioden der Kamera her, bei den beiden anderen Objekten dagegen von der Vorwärts- bzw. Rückwärtsbewegung der Kamera.

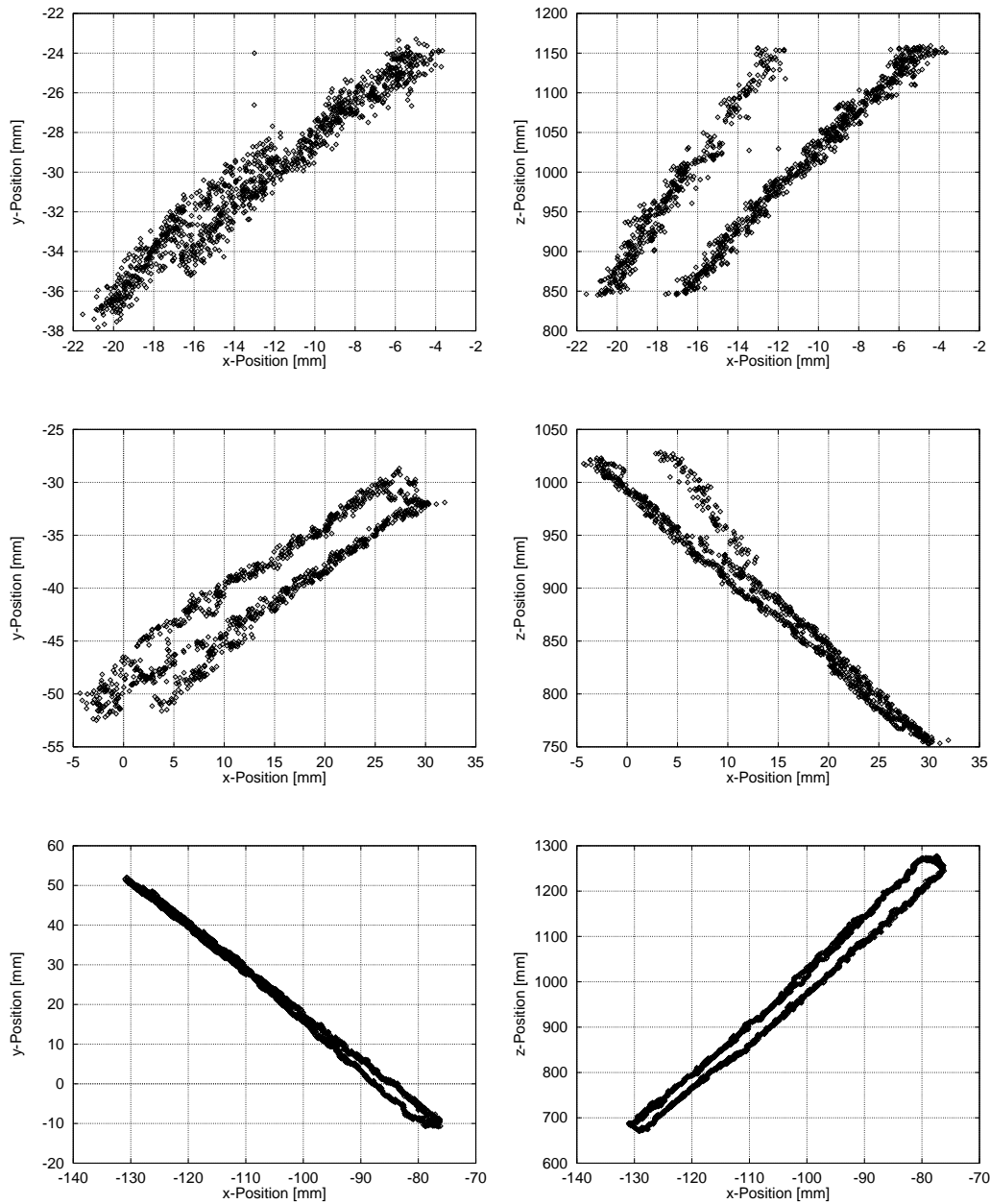


Abbildung 5.1: Positionsschätzungen $(x, y, z)^T$ bei Verfolgung der Batterie (oben), des Luftfilters (Mitte) sowie des kugelförmigen Behälters (unten) bezüglich des bewegten Kamerakoordinatensystems während zwei Bewegungsperioden der Kamera. Zu Darstellungszwecken wurden die Positionsschätzungen jeweils in die xy - (links) und xz -Ebene (rechts) projiziert.

Kapitel 6

Zusammenfassung und Ausblick

Anhand der beiden gewählten unterschiedlichen Teilbereiche innerhalb des Feldes modellbasierter Sichtsysteme konnte herausgearbeitet werden, wie unterschiedliche Rahmenbedingungen und Anforderungen zu jeweils spezifischen Priorisierungen der einzelnen Bewertungsaspekte führen. Es wurde dargestellt, welche Bewertungsaspekte hierbei in welchen Situationen auftreten und wie sie als natürlicher Teil des Entwicklungsprozesses modellbasierter Systeme entstehen. Die sich herauskristallisierenden Fragestellungen hinsichtlich Automatisierbarkeit, Genauigkeit, Robustheit, Zuverlässigkeit, Sicherheit, Ausfallsicherheit, Flexibilität, Ausführungsgeschwindigkeit, sowie Analysierbarkeit wurden erläutert und zu allen Aspekten Lösungsansätze vorgestellt und umgesetzt.

Bei der Behandlung der einzelnen Aspekte ergab sich als herauszustellende Erkenntnis, daß hierzu zwei Möglichkeiten zur Erzielung eines signifikanten Fortschrittes genutzt werden müssen. Im einen Fall handelt es sich um die Entwicklung neuer Algorithmen und Verfahren, im zweiten um die Verbesserung zahlreicher Einzelfaktoren. Insbesondere letzteres erweist sich bei einigen Fragestellungen als unumgänglich, und so wurden beide Möglichkeiten auch genutzt.

Zusammenfassend wurden in der vorliegenden Arbeit folgende Ergebnisse erzielt:

- Als eines der wohl wichtigsten Ergebnisse ist die Realisierung der automatischen, *realwelttauglichen*, modellbasierten *mono- oder multiokularen Objektlokalisierung* zur Initialisierung des für die sichtsystemgestützte Regelung erforderlichen Verfolgungsprozesses zu nennen. Damit konnte die *Automatisierbarkeit* durch eine signifikante *Senkung des Interaktionsgrades* bei der Demontage deutlich verbessert und die Kette zwischen Bildeinzug und Regelung geschlossen werden. Die wesentlichen Merkmale der Lokalisation lassen sich wie folgt stichwortartig zusammenfassen:
 - Hierarchisch organisierte Abtastung im Szenenraum, d.h. im Raum zugelassener Objektlagen (bis 6D).

- Bewußt keine lokale Suchmethode, sondern *globale Suche*, die aus Rechenzeitgründen eingeschränkt werden kann, dadurch weitgehender Ausschluß lokaler Optima.
 - Bewertung der abgetasteten Lagen durch ein von der Granularität der Suchstufe abhängiges Gütemaß.
 - Durch sukzessive Verkleinerung der Granularität verringern sich die Kantensuchbereiche, d.h. es handelt sich um ein selbst zuziehendes Verfahren.
 - Kombination der Suche im Szenenraum mit einer Variante der Hough-Transformation über die Gütefunktion.
 - Gütefunktion nicht differenzierbar oder ableitbar.
 - Explizite Berücksichtigung von möglichen Fehlzuordnungen zwischen projiziertem Modell und Bilddaten.
 - Annahme, daß *eine der* am besten bewerteten Modellagen innerhalb einer Suchstufe in weiteren Iterationen zu der richtigen Modellage führt. Die endgültige Entscheidung wird in der letzten Iteration getroffen.
 - Einbeziehung von Position und Orientierung der Kantenelemente. Es wird explizit berücksichtigt, daß die extrahierte Richtung im Kantenelementbild bei geraden, realen Bildkanten um bis zu $\pm 6^\circ$ variiert.
 - A-priori Wissen über die Objektlage kann einfach berücksichtigt werden.
 - Die Lokalisation funktioniert mono- und binokular.
 - Abbildung von Anpassungsproblemen, eventuellen Zuordnungsmehrdeutigkeiten und Problemen mit komplexem Hintergrund auf ein Rechenzeitproblem.
 - Das Verfahren ist robust (bzgl. aller Testbeispiele).
 - Einziger bislang identifizierter Nachteil: ohne weitere Maßnahmen hohe Rechenzeiten (wenige Minuten bis 46 Minuten).
- Eine Verbesserung der *Automatisierung* wurde bei der Personenverfolgung erzielt durch die *Senkung des Interaktionsgrades* mittels Übergang von einer bildfolgenindividuellen zu einer *bildfolgenunabhängigen Parametrisierung* des Verfahrens. Dabei konnte die im Mittel erreichte *Verfolgungsgüte verbessert* werden.
 - Des Weiteren ist die deutliche Steigerung von *Robustheit* bzw. *Anpassungsgenauigkeit* durch die Verwendung bzw. Verbesserung einer *adaptiven Bildfilterung* bei der Demontage zu nennen sowie der Einsatz eines *iterativ erweiterten Kalman-Filters unter Verwendung einer Levenberg-Marquardt-Optimierung*

bei der Personenverfolgung (anstatt eines erweiterten Kalmanfilters mit Gauß-Newton-Optimierung) und die systematische Verbesserung mehrerer Einzelfaktoren.

- Die *Flexibilität* konnte bei der Demontage verbessert werden durch eine Steigerung der Geräte- sowie Bildformatunabhängigkeit und der Erweiterung um die Möglichkeit, beliebige *Bahnen* mit den *Beobachtungskameras* abfahren zu können. Zudem wurde das Einsatzgebiet erweitert um die *Verfolgung zweier neuer Objekte*, die das Verfahren vor weitere Herausforderungen stellen (kugelförmiger Behälter, Schiffsteil), sowie um die *Regelung des Schrauberwerkzeuges* zum erfolgreichen Lösen spezieller Schraubverbindungen (Realisation der Schrauberregelung nach [Gengenbach 94] mittels der Routinen in *Visdom*).
- Ferner ist besonders auf die *Beschleunigung* der Verfolgung bei der Demontage unter Durchführung verschiedener Maßnahmen hinzuweisen, was nun eine *Regelung in Echtzeit* ermöglicht. Aus Praktikabilitätsgründen wurde zudem die (robuste, aber dadurch rechenintensive) Lokalisation durch eine *verteilte Ausführung auf mehreren Rechnern bzw. Prozessoren* um ein Vielfaches beschleunigt.
- *Zuverlässigkeit, Sicherheit* und *Ausfallsicherheit* konnten verbessert werden durch Verwendung eines neuen Steuerrechners, einer Reorganisation der Gerätestruktur, einer Reimplementation von Bildeinzug und Roboteransteuerung sowie vor allem durch eine *systematische Bereinigung des historisch gewachsenen Programmsystems* (was sich auch auf weitere Aspekte positiv auswirkte).
- Aufgrund des komplexen Zusammenspiels zahlreicher, meist nicht überschaubarer Einzeleffekte kommt der *Analysierbarkeit* von Ergebnissen wie auch von auftretenden Fehlern eine besondere Bedeutung zu. Hinsichtlich dieses Aspekts konnten Verbesserungen erreicht werden durch sämtliche im vorausgehenden Punkt genannten Maßnahmen sowie durch eine Verbesserung der Benutzungsschnittstelle bzw. -oberfläche.
- Die beiden Teilgebiete wurden einander gegenübergestellt, um aufzuzeigen, in welchen Situationen bei modellbasierten Sichtsystemen welche Aspekte als natürlicher Bestandteil der Entwicklung im Vordergrund stehen. Durch Ableitung allgemeingültiger Aussagen konnte insbesondere eine Parallele zu dynamischen Systemen herausgearbeitet und verdeutlicht werden.

Die durchgeführten Maßnahmen decken sowohl theoretische als auch praktische Überlegungen und Ansätze ab, sowie Geräte-, Programm- und Entwurfsebene. Neben der Entwicklung neuer Verfahren wurde auch der Fortschritt mittels der Verbesserung zahlreicher Einflußfaktoren erläutert. Dies verdeutlicht den Handlungsspielraum, der bei modellbasierten Sichtsystemen zur Verfügung steht.

Als Bewertungsmaßstab für einige der Ergebnisse sei beispielsweise die nun automatisch initialisierte Entnahme der Batterie aus dem Motorraums erwähnt. Im Gegensatz

zu dem vor dieser Arbeit bestehenden System funktioniert sie nun bei praktisch jedem Lauf einwandfrei. Zudem kann nun die Regelung schnell auf äußere Gegebenheiten reagieren, da die monokulare Verarbeitung *schritthaltend mit dem Bildeinzug* funktioniert und die binokulare nur noch bis auf einen Faktor zwei davon entfernt ist.

Wie bei (fast) allen Forschungen kann natürlich auch die vorliegende Arbeit nicht für sich beanspruchen, alle Probleme gelöst zu haben. Neben der Vorstellung der Fortschritte sind daher auch die verbleibenden Schwierigkeiten zu nennen, für die noch keine Lösungen gefunden bzw. die aus Zeitgründen nicht mehr bearbeitet werden konnten. Als Ausblick für weitere Untersuchungen und Verbesserungen sei daher hingewiesen auf die in den Abschnitten 3.1.5 und 4.2.2 dokumentierten verbleibenden Robustheitsschwächen bei der Personenverfolgung bzw. der Demontage (siehe hierzu auch Abschnitt 3.1.4). Zur Vermeidung der in Abschnitt 4.2.2 erwähnten Abhängigkeit der Regelung von starken Beleuchtungsänderungen ist an den Einsatz von Hochdynamikkameras zu denken. Zur Steigerung der Robustheit kann aber vor allem die Gütefunktion der Lokalisation als Kriterium zur Verifikation der Modellanpassung bei der Verfolgung herangezogen werden. Falls damit festgestellt werden kann, wann eine Verfolgung nicht mehr genau funktioniert, kann die Verfolgung neu aufgesetzt werden und so das Problem eines zu kleinen Fangbereichs im Bild umgangen werden, vgl. Abschnitt 3.1.5. Das gilt sowohl für den Bereich der Personenverfolgung als auch den Bereich der Roboterregelung. Bei letzterem sichert dies insbesondere die Robustheit und Sicherheit des Systems. Ferner kann bei der Demontage versucht werden, auch bei der binokularen Verfolgung bzw. Regelung *jedes* Halbbild schritthaltend auszuwerten. Bisher scheitert dies hauptsächlich an der Synchronisation der Stereobilder.

In Zukunft können durch den erreichten Stand zudem immer mehr Erweiterungen der Teilbereiche ins Auge gefaßt werden. Man denke etwa an die Nachführung der Kamera bei der Personenverfolgung oder die Erfassung weiterer Bewegungsvorgänge von Personen. Bei der Demontage kann überlegt werden, wie die verschiedenen Komponenten des Motorraums gelöst und anschließend entnommen werden können. Zum Lösen der Verbindungen könnte dabei beispielsweise ein Wasserstrahlschneider zum Einsatz kommen.

Literaturverzeichnis

- [Alter & Jacobs 98] T.D. Alter, D.W. Jacobs: *Uncertainty Propagation in Model-Based Recognition*. International Journal of Computer Vision **27:2** (1998) 127–159.
- [Bar-Shalom & Fortmann 88] Y. Bar-Shalom, T. E. Fortmann: *Tracking and Data Association*. Academic Press, San Diego, CA, USA, 1988.
- [Basri & Jacobs 99] R. Basri, D. Jacobs: *Projective Alignment with Regions*. Proc. 7th International Conference on Computer Vision ICCV '99, 20–27 September 1999, Kerkyra, Greece, pp. 20–27.
- [Batlle et al. 00] J. Batlle, A. Casals, J. Freixenet, J. Martí: *A Review on Strategies for Recognizing Natural Objects in Colour Images of Outdoor Scenes*. Image and Vision Computing **18:6-7** (2000) 515–530.
- [Bebis et al. 98] G. Bebis, M. Georgiopoulos, M. Shah, N.d.V. Lobo: *Using Algebraic Functions of Views for Indexing-Based Object Recognition*. Proc. 6th International Conference on Computer Vision ICCV '98, 4–7 January 1998, Bombay, India, pp. 634–639.
- [Beis & Lowe 99] J.S. Beis, D.G. Lowe: *Indexing without Invariants in 3D Object Recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-**21:10** (1999) 1000–1015.
- [Benameur & Bélanger 98] K. Benameur, P.R. Bélanger: *Grasping of a Moving Object with a Robotic Hand-Eye System*. Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems, October 1998, Victoria, B.C., Canada, pp. 304–310.
- [Black & Jepson 98] M.J. Black, A.D. Jepson: *EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation*. International Journal of Computer Vision **26:1** (1998) 63–84.
- [Boehm & Prautzsch 93] W. Boehm, H. Prautzsch: *Numerical Methods*. Vieweg-Verlag, Braunschweig/Wiesbaden 1993.
- [Borotschnig et al. 00] H. Borotschnig, L. Paletta, M. Prantl, A. Pinz: *Appearance-Based Active Object Recognition*. Image and Vision Computing **18:9** (2000) 715–727.
- [Boykov & Huttenlocher 99] Y. Boykov, D.P. Huttenlocher: *A New Bayesian Frame-*

- work for Object Recognition*. Proc. IEEE Computer Vision and Pattern Recognition CVPR '99, 23–25 June 1999, Fort Collins, Colorado, Vol. II, pp. 517–523.
- [Cass 98] T.A. Cass: *Robust Affine Structure Matching for 3D Object Recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-**20**:11 (1998) 1265–1274.
- [Chang & Tsai 99] C.-C. Chang, W.-H. Tsai: *Reliable Determination of Object Pose from Line Features by Hypothesis Testing*. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-**21**:12 (1999) 1235–1241.
- [Chau & Siu 99] C.-P. Chau, W.-C. Siu: *Generalized Dual-Point Hough Transform for Object Recognition*. Proc. International Conference on Image Processing ICIP '99, 24–28 October, 1999, Kobe, Japan, Vol. I, pp. 560–564.
- [Chen & Stockman 98] J.-L. Chen, G.C. Stockman: *3D Free-Form Object Recognition Using Indexing by Contour Features*. Computer Vision and Image Understanding **71**:3 (1998) 334–355.
- [Chhabra & Phillips 98] A.K. Chhabra, I.T. Phillips: *A Benchmark for Graphics Recognition System*. In 'Empirical Evaluation Techniques in Computer Vision', K.W. Bowyer, P.J. Phillips (Eds.), IEEE Computer Society Press, Los Alamitos/CA, Washington, Brussels, Tokyo, 1998, pp. 28–38.
- [Crétual et al. 98] A. Crétual, F. Chaumette, P. Bouthemy: *Complex Object Tracking by Visual Servoing Based on 2D Image Motion*. Proc. IEEE 14th International Conference on Pattern Recognition, 16–20 August 1998, Brisbane, Australia, Vol. 2, pp. 1251–1254.
- [Daniilidis 92] K. Daniilidis: *Zur Fehlerempfindlichkeit in der Ermittlung von Objektbeschreibungen und relativen Bewegungen aus monokularen Bildfolgen*. Dissertation, Fakultät für Informatik der Universität Karlsruhe (TH), Juli 1992.
- [Das & Bhanu 98] S. Das, B. Bhanu: *A System for Model-Based Object Recognition in Perspective Aerial Images*. Pattern Recognition **31**:4 (1998) 465–491.
- [Delamarre & Faugeras 99] Q. Delamarre, O. Faugeras: *3D Articulated Models and Multi-View Tracking with Silhouettes*. Proc. 7th International Conference on Computer Vision ICCV '99, 20–27 September 1999, Kerkyra, Greece, pp. 716–721.
- [Drummond & Cipolla 00] T. Drummond, R. Cipolla: *Real-Time Tracking of Multiple Articulated Structures in Multiple Views*. 6th European Conference on Computer Vision ECCV-2000, June/July 2000, Dublin, Ireland, Lecture Notes in Computer Science 1843, D. Vernon (Ed.), Proc. Vol. II, Springer-Verlag: Berlin, Heidelberg, New York/NY, pp. 20–36.
- [Engeln-Müllges & Reutter 90] G. Engeln-Müllges, F. Reutter: *Formelsammlung zur Numerischen Mathematik mit C-Programmen*. B.I. Wissenschaftsverlag Mannheim/Wien/Zürich, 2. Auflage 1990.

- [Farin 94] G. Farin: *Kurven und Flächen im Computer Aided Geometric Design. Eine praktische Einführung*. Vieweg Verlag Braunschweig/Wiesbaden, ISBN 3-528-16542-1, 2. Auflage, 1994.
- [Geeter et al. 97] J.D. Geeter, H.V. Brussel, J.D. Schutter: *A Smoothly Constrained Kalman Filter*. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-**19**:10 (1997) 1171–1177.
- [Gelb et al. 74] A. Gelb, J. F. Kasper, R. A. Nash, C. F. Price, A. A. Sutherland: *Applied Optimal Estimation*. MIT Press, Cambridge, MA, USA, 1974.
- [Gengenbach 94] V. Gengenbach: *Einsatz von Rückkopplungen in der Bildauswertung bei einem Hand-Auge-System zur automatischen Demontage*. Dissertation, Fakultät für Informatik der Universität Karlsruhe (TH), Juli 1994; erschienen in Dissertationen zur künstlichen Intelligenz (DISKI), Band **72**, infix-Verlag, Sankt Augustin 1994.
- [Gengenbach et al. 96] V. Gengenbach, H.-H. Nagel, M. Tonko, and K. Schäfer: *Automatic Dismantling Integrating Optical Flow into a Machine-Vision Controlled Robot System*. Proc. IEEE International Conference on Robotics and Automation, 22–28 April 1996, Minneapolis/MN, N. Caplan (Ed.), IEEE Computer Society Press, Los Alamitos/CA, 1996, pp. 1320–1325.
- [Gold et al. 98] S. Gold, A. Rangarajan, C.-P. Lu, S. Pappu, E. Mjolsness: *New Algorithms for 2D and 3D Point Matching: Pose Estimation and Correspondence*. Pattern Recognition **31**:8 (1998) 1019–1031.
- [Grewal/Andrews 93] M. S. Grewal, A. P. Andrews: *Kalman Filtering, Theory and Practice*. Prentice Hall, Englewood Cliffs, New Jersey, USA, 1993.
- [Grimson 90] W.E.L. Grimson: *Object Recognition by Computer: The Role of Geometric Constraints* (with contributions from T. Lozano-Pérez and D.P. Huttenlocher). The MIT Press, Cambridge, Massachusetts, London, England, 1990.
- [Guil et al. 99] N. Guil, J.M. Gonzalez-Linares, E.L. Zapata: *Bidimensional Shape Detection Using an Invariant Approach*. Pattern Recognition **32** (1999) 1025–1038.
- [Haag 98] M. Haag: *Bildfolgenauswertung zur Erkennung der Absichten von Straßenverkehrsteilnehmern*. Dissertation, Fakultät für Informatik der Universität Karlsruhe (TH), Juli 1998; erschienen in Dissertationen zur künstlichen Intelligenz (DISKI), Band **193**, infix-Verlag, Sankt Augustin 1998.
- [Hall et al 00] D. Hall, V.C. Verdière, J.L. Crowley: *Object Recognition Using Coloured Receptive Fields*. Proc. 6th European Conference on Computer Vision ECCV-2000, June/July 2000, Dublin, Ireland, Lecture Notes in Computer Science 1842, D. Vernon (Ed.), Vol. I, Springer-Verlag: Berlin Heidelberg New York/NY 2000, pp. 164–177.

- [Han et al. 99] I. Han, I.D. Yun, S.U. Lee: *Model-Based Object Recognition Using the Hausdorff Distance with Explicit Pairing*. International Conference on Image Processing ICIP '99, 24–28 October 1999, Kobe, Japan, Vol. IV, pp. 83–87.
- [Häusler & Ritter 99] G. Häusler, D. Ritter: *Feature-Based Object Recognition and Localization in 3D-Space, Using a Single Video Image*. Computer Vision and Image Understanding **73**:1 (1999) 64–81.
- [Heimes 00] F. Heimes: *Sichtsystemgestützte Fahrerassistenz im innerstädtischen Straßenverkehr*. Dissertation, Fakultät für Informatik der Universität Karlsruhe (TH), 20. Juli 2000; erschienen in Dissertationen zur künstlichen Intelligenz (DISKI), Band **243**, Akademische Verlagsgesellschaft, Berlin, 2000.
- [Herbin 98] S. Herbin: *Combining Geometric and Probabilistic Structure for Active Recognition of 3D Objects*. 5th European Conference on Computer Vision ECCV '98, Freiburg, Germany, June 1998, Lecture Notes in Computer Science 1407, H. Burkhardt, B. Neumann (Eds.), Vol. II, Springer-Verlag Berlin, Heidelberg, pp. 748–764.
- [Horaud & Sossa 95] R. Horaud, H. Sossa: *Polyhedral Object Recognition by Indexing*. Pattern Recognition **28**:12 (1995) 1855–1870.
- [Hsu & Aquino 99] L. Hsu, P.L.S. Aquino: *Adaptive Visual Tracking with Uncertain Manipulator Dynamics and Uncalibrated Cameras*. Proc. IEEE 38th Conference on Decision and Control, December 1999, Phoenix, Arizona, USA, pp. 1248–1253.
- [Huang et al. 00] C.-L. Huang, M.-S. Wu, S.-H. Jeng: *Gesture Recognition Using the Multi-PDM Method and Hidden Markov Model*. Image and Vision Computing **18**:11 (2000) 865–879.
- [Huttenlocher et al. 93] D.P. Huttenlocher, G.A. Klanderman, W.J. Rucklidge: *Comparing Images Using the Hausdorff Distance*. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-**15**:9 (1993) 850–863.
- [Huttenlocher et al. 99] D.P. Huttenlocher, R.H. Lilien, C.F. Olson: *View-Based Recognition Using an Eigenspace Approximation to the Hausdorff Measure*. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-**21**:9 (1999) 951–955.
- [Johnson & Hebert 99] A.E. Johnson, M. Hebert: *Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes*. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-**21**:5 (1999) 433–449.
- [Joseph 99] S.H. Joseph: *Analysing and Reducing the Cost of Exhaustive Correspondence Search*. Image and Vision Computing **17**:11 (1999) 815–830.
- [Jurie 98a] F. Jurie: *Hypothesis Verification in Model-Based Object Recognition with a Gaussian Error Model*. Proc. 5th European Conference on Computer

- Vision ECCV '98, Freiburg, Germany, June 1998, Lecture Notes in Computer Science 1407, H. Burkhardt, B. Neumann (Eds.), Vol. II, Springer-Verlag Berlin, Heidelberg, pp. 642–656.
- [Jurie 98b] F. Jurie: *Tracking Objects with a Recognition Algorithm*. Pattern Recognition Letters **19**:3-4 (1998) 331–340.
- [Jurie 99] F. Jurie: *Robust Hypothesis Verification: Application to Model-Based Object Recognition*. Pattern Recognition **32** (1999) 1069–1081.
- [Kakadiaris & Metaxas 00] I. Kakadiaris, D. Metaxas: *Model-Based Estimation on 3D Human Motion*. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-**22**:12 (2000) 1453–1459.
- [Kassim et al. 99] A.A. Kassim, T. Tan, K.H. Tan: *A Comparative Study of Efficient Generalized Hough Transform Techniques*. Image and Vision Computing **17**:10 (1999) 737–748.
- [Kiryati et al. 00] N. Kiryati, H. Kälviäinen, S. Alaoutinen: *Randomized or probabilistic Hough Transform: Unified Performance Evaluation*. Pattern Recognition Letters **21** (2000) 1157–1164.
- [Koller 92] D. Koller: *Detektion, Verfolgung und Klassifikation bewegter Objekte in monokularen Bildfolgen am Beispiel von Straßenverkehrsszenen*. Dissertation, Fakultät für Informatik der Universität Karlsruhe (TH), 1992; erschienen in Dissertationen zur künstlichen Intelligenz (DISKI), Band **13**, infix-Verlag, Sankt Augustin 1992.
- [Lee et al. 99] S.-H. Lee, H.-K. Hong, J.-S. Choi: *A Study on Assembly Part Recognition Using Part-Based Superquadric Model*. International Conference on Image Processing ICIP '99, 24–28 October 1999, Kobe, Japan, Vol. IV, pp. 78–82.
- [Lee & Bailey-Van Kuren 00] K.-M. Lee, M.M. Bailey-Van Kuren: *Modeling and Supervisory Control of a Disassembly Automation Workcell Based on Blocking Topology*. IEEE Transactions on Robotics and Automation **16**:1 (2000) 67–77.
- [Lengyel 98] J. Lengyel: *The Convergence of Graphics and Vision*. IEEE Computer **31**:7 (1998) 46–53.
- [Leuck 00] H. Leuck: *Untersuchungen zu einer systematischen Leistungssteigerung in der modellbasierten Bildauswertung*. Dissertation, Fakultät für Informatik der Universität Karlsruhe (TH), 27. Juli 2000.
- [Li et al. 98] Z. Li, J. Gou, Y. Chu: *Geometric Algorithms for Workpiece Localization*. IEEE Transactions on Robotics and Automation **14**:6 (1998) 864–878.
- [Lo & Tsai 97] R.-C. Lo, W.-H. Tsai: *Perspective-Transformation-Invariant Generalized Hough Transform for Perspective Planar Shape Detection and Matching*. Pattern Recognition **30**:3 (1997) 383–396.
- [Lowe 99] D.G. Lowe: *Object Recognition from Local Scale-Invariant Features*. Proc. 7th IEEE International Conference on Computer Vision ICCV '99,

- 20–27 September 1999, Kerkyra, Greece, Vol. II, 1999, pp. 1150–1157.
- [Lu et al. 00] C.-P. Lu, G.D. Hager, E. Mjølness: *Fast and Globally Convergent Pose Estimation from Video Images*. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-22:6 (2000) 610–622.
- [Marchand et al. 99] É. Marchand, P. Bouthemy, F. Chaumette, V. Moreau: *Robust Visual Tracking by Coupling 2D Motion and 3D Pose Estimation*. International Conference on Image Processing ICIP '99, 24–28 October 1999, Kobe, Japan, Vol. IV, pp. 98–102.
- [Mirmehdi et al. 99] M. Mirmehdi, P.L. Palmer, J. Kittler, H. Dabis: *Feedback Control Strategies for Object Recognition*. IEEE Transactions on Image Processing 8:8 (1999) 1084–1101.
- [Müller & Nagel 99] Th. Müller, H.-H. Nagel: *Untersuchungen zur sichtsystemgestützten automatischen Demontage*. Interner Bericht I, Institut für Algorithmen und Kognitive Systeme, Fakultät für Informatik der Universität Karlsruhe (TH), 31. August 1999.
- [Müller & Nagel 00a] Th. Müller, H.-H. Nagel: *Untersuchungen zur sichtsystemgestützten automatischen Demontage*. Interner Bericht II, Institut für Algorithmen und Kognitive Systeme, Fakultät für Informatik der Universität Karlsruhe (TH), 9. Januar 2001..
- [Müller & Nagel 00b] Th. Müller, H.-H. Nagel: *Untersuchungen zur Verfolgung mehrgliedriger Objekte in Bildfolgen*. Interner Bericht, Institut für Algorithmen und Kognitive Systeme, Fakultät für Informatik der Universität Karlsruhe (TH), 8. Januar 2001.
- [Murase & Nayar 95] H. Murase, S.K. Nayar: *Visual Learning and Recognition of 3-D Objects from Appearance*. International Journal of Computer Vision 14:1 (1995) 5–24.
- [Nagel & Gehrke 98] H.-H. Nagel, A. Gehrke: *Spatiotemporally Adaptive Estimation and Segmentation of OF-Fields*. 5th European Conference on Computer Vision ECCV '98, 2–6 June 1998, Freiburg, Proceedings Vol. II; Lecture Notes in Computer Science 1407, H. Burkhardt, B. Neumann (Eds.), Springer-Verlag Berlin Heidelberg 1998, pp. 86–102.
- [Nagel et al. 99] H.-H. Nagel, T. Müller, V. Gengenbach, A. Bachem, H. Leuck: *Machine Vision Competence as a Function of Computing Power – 4½ Years Later*. European Control Conference ECC '99, 31 August–3 September 1999, Karlsruhe/Germany.
- [Olson 98] C.F. Olson: *A Probabilistic Formulation for Hausdorff Matching*. Proc. IEEE Conference on Computer Vision and Pattern Recognition CVPR '98, 23–25 June 1998, Santa Barbara, California, pp. 150–156.
- [Olson 99] C.F. Olson: *A General Method for Feature Matching and Model Extraction*. Proc. Workshop on 'Vision Algorithms – Theory and Practice', 21–22 September 1999, Corfu, Greece (in conjunction with ICCV '99),

- Lecture Notes in Computer Science 1889, B. Triggs, R. Szeliski, A. Zisserman (Eds.), Springer-Verlag Berlin Heidelberg 2000, pp. 20-26.
- [Olson 00] C.F. Olson: *Probabilistic Self-Localization for Mobile Robots*. IEEE Transactions on Robotics and Automation **16**:1 (2000) 55–66.
- [Otte 94] M. Otte: *Extraktion von linienförmigen Merkmalen und Ermittlung des optischen Flusses mit seinen Ableitungen aus Voll- und Halbbildfolgen*. Dissertation, Fakultät für Informatik der Universität Karlsruhe (TH), Februar 1994; erschienen in Dissertationen zur künstlichen Intelligenz (DISKI), Band **63**, infix-Verlag, Sankt Augustin 1994.
- [Paulik & Wang 98] M.J. Paulik, Y.D. Wang: *Three-Dimensional Object Recognition Using Vector Wavelets*. Proc. International Conference on Image Processing ICIP '98, 4–7 October 1998, Chicago, Illinois, USA, pp. 586–590.
- [Porill et al. 88] J. Porill, S.B. Pollard, T.P. Pridmore, J.B. Bowen, J.E.W. Mayhew, J.P. Frisby: *TINA: A 3D Vision System for Pick and Place*. Image and Vision Computing **6**:2 (1988) 91–99.
- [Press et al. 92] W. H. Press, W. T. Vetterling, S. A. Teukolsky, B. P. Flannery: *Numerical Recipes in C, The Art of Scientific Computing*. Cambridge University Press, Cambridge, UK, 2. Auflage 1992.
- [Ratan et al. 98] A.L. Ratan, W.E.L. Grimson, W.M. Wells: *Object Detection and Localization by Dynamic Template Warping*. Proc. IEEE Conference on Computer Vision and Pattern Recognition CVPR '98, 23–25 June 1998, Santa Barbara, California, pp. 634–640.
- [Ratan et al. 00] A.L. Ratan, W.E.L. Grimson, W.M. Wells: *Object Detection and Localization by Dynamic Template Warping*. International Journal of Computer Vision **36**:2 (2000) 131–147.
- [Rechtin 97] E. Rechtin: *The Synthesis of Complex Systems*. IEEE Spectrum **34**:7 (1997) 51–55.
- [Rosin 99] P.L. Rosin: *Robust Pose Estimation*. IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics **29**:2 (1999) 297–303.
- [Ruf & Horaud 00] A. Ruf, R. Horaud: *Vision-Based Guidance and Control of Robots in Projective Space*. 6th European Conference on Computer Vision ECCV-2000, June/July 2000, Dublin, Ireland, Lecture Notes in Computer Science 1843, D. Vernon (Ed.), Proc. Vol. II, Springer-Verlag: Berlin, Heidelberg, New York/NY, pp. 50–66.
- [Schäfer 96] K. Schäfer: *PATTY, Preprocessor for Attributed Types*. Interner Bericht, Institut für Algorithmen und Kognitive Systeme, Fakultät für Informatik der Universität Karlsruhe (TH), 1996.
- [Sclaroff & Pentland 95] S. Sclaroff, A.P. Pentland: *Modal Matching for Correspondence and Recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-**17**:6 (1995) 545–561.

- [Sidenbladh et al. 00a] H. Sidenbladh, F. De la Torre, M.J. Black: *A Framework for Modeling the Appearance of 3D Articulated Figures*. Proc. 4th IEEE International Conference on Automatic Face and Gesture Recognition, 28–30 March 2000, Grenoble, France, pp. 368–375.
- [Sidenbladh et al. 00b] H. Sidenbladh, M.J. Black, D.J. Fleet: *Stochastic Tracking of 3D Human Figures Using 2D Image Motion*. 6th European Conference on Computer Vision ECCV-2000, June/July 2000, Dublin, Ireland, Lecture Notes in Computer Science 1843, D. Vernon (Ed.), Proc. Vol. II, Springer-Verlag: Berlin Heidelberg New York/NY, pp. 702–718.
- [Silberberg et al. 84] T.M. Silberberg, L. Davis, D. Harwood: *An Iterative Hough Procedure for Three-Dimensional Object Recognition*. Pattern Recognition **17:6** (1984) 621–629.
- [Sim et al. 99] D.-G. Sim, O.-K. Kwon, R.-H. Park: *Object Matching Algorithms Using Robust Hausdorff Distance Measures*. IEEE Transactions on Image Processing **8:3** (1999) 425–429.
- [Sullivan et al. 99] J. Sullivan, A. Blake, M. Isard, J. MacCormick: *Object Localization by Bayesian Correlation*. Proc. IEEE 7th International Conference on Computer Vision ICCV '99, 20–27 September 1999, Kerkyra, Greece, Vol. II, pp. 1068–1075.
- [Sullivan et al. 00] J. Sullivan, A. Blake, J. Rittscher: *Statistical Foreground Modelling for Object Localisation*. 6th European Conference on Computer Vision ECCV-2000, June/July 2000, Dublin, Ireland, Lecture Notes in Computer Science 1843, D. Vernon (Ed.), Proc. Vol. II, Springer-Verlag: Berlin Heidelberg New York/NY, pp. 307–323.
- [Teubner 96] Teubner: *Taschenbuch der Mathematik*. Teubner Stuttgart, Leipzig 1996.
- [Tonko 97] M. Tonko: *Zur sichtsystemgestützten Demontage am Beispiel von Altfahrzeugen*. Dissertation, Fakultät für Informatik der Universität Karlsruhe (TH), Juni 1997; erschienen in Dissertationen zur künstlichen Intelligenz (DISKI), Band **166**, infix-Verlag, Sankt Augustin 1997.
- [Toth & Culverhouse 99] L. Toth, P.F. Culverhouse: *3D Object Recognition from Static 2D Views Using Multiple Coarse Data Channels*. Image and Vision Computing **17:11** (1999) 845–858.
- [Vel & Aeberhard 00] O. Vel, S. Aeberhard: *Object Recognition Using Random Image Lines*. Image and Vision Computing **18:3** (2000) 193–198.
- [Vijayakumar et al. 98] B. Vijayakumar, D. Kriegman, J. Ponce: *Invariant-Based Recognition of Complex Curved 3D Objects from Image Contours*. Computer Vision and Image Understanding **72:3** (1998) 287–303.
- [Wachter 97] S. Wachter: *Verfolgung von Personen in monokularen Bildfolgen*. Dissertation, Fakultät für Informatik der Universität Karlsruhe (TH), Juli 1997; erschienen im Vice Versa Verlag, Berlin 1997.

- [Wachter & Nagel 99] S. Wachter, H.-H. Nagel: *Tracking Persons in Monocular Image Sequences*. Computer Vision and Image Understanding **74:3** (1999) 174–192.
- [Wang 99] P.S.P. Wang: *High Level Visualization, Representation, Understanding, and Recognition of 3D Articulated Objects*. The Encyclopedia of Microcomputers, A. Kent, J. Williams (Eds.), Marcer Dekker Pub. Co., 1999 (elektronisch verfügbar unter <http://www.ccs.neu.edu/home/pwang/publications.html>).
- [Weiss & Ray 98] I. Weiss, M. Ray: *Model-Based Recognition of 3D Objects from One View*. 5th European Conference on Computer Vision ECCV '98, June 1998, Freiburg, Germany, Lecture Notes in Computer Science 1407, H. Burkhardt, B. Neumann (Eds.), Vol. II, Springer-Verlag Berlin, Heidelberg, pp. 716–732.
- [Wheeler & Ikeuchi 95] M.D. Wheeler, K. Ikeuchi: *Sensor Modeling, Probabilistic Hypothesis Generation, and Robust Localization for Object Recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-**17:3** (1995) 252–265.
- [Wiedemann et al. 98] C. Wiedemann, C. Heipke, H. Mayer: *Empirical Evaluation of Automatically Extracted Road Axes*. In 'Empirical Evaluation Techniques in Computer Vision', K.W. Bowyer, P.J. Phillips (Eds.), IEEE Computer Society Press, Los Alamitos/CA, Washington, Brussels, Tokyo, 1998, pp. 28–38.
- [Winkeler et al. 99] J. Winkeler, B.S. Manjunath, S. Chandrasekaran: *Subset Selection for Active Object Recognition*. Proc. IEEE Computer Vision and Pattern Recognition CVPR '99, 23–25 June 1999, Fort Collins, Colorado, pp. 511–516.
- [Worthington & Hancock 00] P.L. Worthington, E.R. Hancock: *Region-Based Object Recognition Using Shape-from-Shading*. 6th European Conference on Computer Vision ECCV-2000, June/July 2000, Dublin, Ireland, Lecture Notes in Computer Science 1842, D. Vernon (Ed.), Proc. Vol. I, Springer-Verlag: Berlin Heidelberg New York/NY 2000, pp. 455–471.
- [Wren et al. 97] C.R. Wren, A. Azarbayejani, T. Darrell, A.P. Pentland: *Pfinder: Real-Time Tracking of the Human Body*. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-**19:7** (1997) 780–785.
- [Yáñez-Suárez & Azimi-Sadjadi 99] O. Yáñez-Suárez, M.R. Azimi-Sadjadi: *Unsupervised Clustering in Hough Space for Identification of Partially Occluded Objects*. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-**21:5** (1999) 946–950.
- [Yang & Cohen 99] Z. Yang, F.S. Cohen: *Image Registration and Object Recognition Using Affine Invariants and Convex Hulls*. IEEE Transactions on Image Processing **8:7** (1999) 934–946.

- [Yang et al. 00] M.-H. Yang, D. Roth, N. Ahuja: *Learning to Recognize 3D Objects with SNoW*. 6th European Conference on Computer Vision ECCV-2000, June/July 2000, Dublin, Ireland, Lecture Notes in Computer Science 1842, D. Vernon (Ed.), Proc. Vol. I, Springer-Verlag Berlin, Heidelberg, New York/NY, pp. 439–454.
- [Yeasin & Chaudhuri 00] M. Yeasin, S. Chaudhuri: *Toward Automatic Robot Programming: Learning Human Skill from Visual Data*. IEEE Transactions on Systems, Man, and Cybernetics–Part B: Cybernetics **30**:1 (2000) 180–185.
- [Zergeroglu et al. 99a] E. Zergeroglu, D.M. Dawson, M.S. Queiroz, S. Nagarkatti: *Robust Visual-Servo Control of Robot Manipulators in the Presence of Uncertainty*. Proc. IEEE 38th Conference on Decision and Control, December 1999, Phoenix, Arizona, USA, pp. 4137–4142.
- [Zergeroglu et al. 99b] E. Zergeroglu, D.M. Dawson, M.S. Queiroz, A. Behal: *Vision-Based Nonlinear Tracking Controllers with Uncertain Robot-Camera Parameters*. Proc. IEEE/ASME International Conference on Advanced Intelligent Mechatronics, 19–23 September 1999, Atlanta, USA, pp. 854–859.

Index

A

Ablaufstruktur der Verfolgung 143
 $abst_{\max}$ 103
Abtastverfahren 94
adaptive Bildfilterung 115
adaptive Bildfilterung: Trennung zwischen unwesentlicher und systematischer Grauwertvariation 120
`alinde` 5, 41, 66
[Alter & Jacobs 98] 18
Analysierbarkeit 9, 90, 92, 124, 125, 127, 128, 153, 160
Anfahrngenauigkeit 114
Anpassungsgenauigkeit 38
Anzahl `Visdom`-Quelldateien 128
Anzeigelupe 90, 153
Aperturereffekt 61
approximierte Hessematrix 48
a-priori Wissen zur Objektlage 105
Aspekte 9
Ausfallsicherheit 90, 92, 124, 125, 126, 128, 155, 160
Autobatterie 7, 99, 115, 123
automatische Berechnung von Θ 39
Automatisierbarkeit 9, 62, 92, 93, 154

B

Bahnkurve des Beobachtungsroboters 136
[Bar-Shalom & Fortmann 88] 47
[Basri & Jacobs 99] 20
[Batlle et al. 00] 29
Batterie 7, 99, 108, 109, 115, 123, 156

Batterie im Motorraum des Versuchsfahrzeuges 134
[Bebis et al. 98] 12, 22
Bedienbarkeit 153
[Beis & Lowe 99] 22
[Benamer & Bélanger 98] 33
Beobachtungsroboter 136
Bereinigungen 126
Bewertung einer projizierten Modellkante 103
Bewertungsaspekte 9
Bézier-Kurve 137
Bildbereich 1, 13, 39, 94
Bildfolgen bei der Personenverfolgung 4
[Black & Jepsen 98] 33
[Boehm & Prautzsch 93] 45
[Borotschnig et al. 00] 25
[Boykov & Huttenlocher 99] 27

C

CAD-Modell 1, 6, 94, 97
[Cass 98] 13
[Chang & Tsai 99] 14
chaotischer Schwingungseffekt 41, 155
[Chau & Siu 99] 23
[Chen & Stockman 98] 22
[Chhabra & Phillips 98] 35, 36
[Crétual et al. 98] 33, 34

D

d_1, d_2 96
[Daniilidis 92] 47
[Das & Bhanu 98] 15
de Casteljau-Algorithmus 139
Definition: Genauigkeit 38

- Definition: Robustheit 40
 Demontage 2, 4, 92
 Detektion 11
 d_{\max} 101
 Doppelklotz 130
 Doppelklotzdemontage 130
 [Drummond & Cipolla 00] 32
 dynamisches System 55, 60, 155
- E**
- Echtzeitaspekt 2, 6, 37, 115, 140, 142,
 147, 155, 160
 Echtzeitklasse 145
 Eigenwerte des GLST 117
 EKF 46
 Empfindlichkeit gegenüber Störungen
 61
 [Engeln-Müllges & Reutter 90] 45
 Ergebnisse, Zusammenfassung 158
 Erkennung 11
- F**
- Fangbereich 41, 61, 161
 [Farin 94] 136, 137
 Fehlertoleranz 105
 Feinsuche 99
 Flächeninformation 3, 7, 66, 84, 86
 Flexibilität 3, 9, 34, 37, 129, 131, 136,
 155, 160
 Fortschritt: zwei Möglichkeiten 9
 frankm 5, 52, 66, 72
 Freiheitsgrade 3, 7, 16
 Freiheitsgradwertebeschränkung 4, 52,
 57, 88
- G**
- G 98
 Gauß-Algorithmus mit Spaltenpivotsu-
 che 45
 Gauß-Newton-Algorithmus 3, 7, 46, 160
 g_braun 5
 [Geeter et al. 97] 60
 [Gelb et al. 74] 47
 genaue Anpassung 39
 Genauigkeit 9, 35, 38, 49, 61, 64, 101,
 106, 114, 124, 140, 160
 Genauigkeit einer Modellanpassung 38
 Genauigkeitsschwelle 39, 114
 [Gengenbach 94] 6, 130, 160
 [Gengenbach et al. 96] 13
 Gerätekonfiguration 142
 Gesamtschrittverfahren 45
 geschlossene Splinekurve 138
 Geschwindigkeitsaspekt 2, 9, 35, 37, 92,
 115, 123, 140, 155
 GIGAmachine 6, 125
 globales Minimum 27, 61, 94, 102, 159
 globales Suchverfahren 94
 GLST-Eigenwerte 117
 $GLST_{\text{init}}(\mathbf{x})$ 116
 [Gold et al. 98] 20
 Gradientenbetragsschwelle 6, 115
 Granularität bei der Objektlokalisierung
 95
 Grauwert-lokaler Strukturtensor 116
 Greifer 4, 7, 9, 92, 114, 123
 Greifermodell 124
 Greiferregelung 114
 [Grewal/Andrews 93] 47
 [Grimson 90] 12, 34
 Grobsuche 99
 Größe von Visdom 128
 Grundwahrheit 39
 [Guil et al. 99] 23
 gute Anpassung 39
 Gütefunktion 46, 61, 77, 86, 94, 95, 98,
 159
 Gütefunktion zur Bewertung einer Mo-
 dellage 94
- H**
- H 48
 [Haag 98] 116
 [Häusler & Ritter 99] 16
 [Hall et al 00] 25

- Handlungsspielraum 160
 [Han et al. 99] 16
 [Heimes 00] 35, 36
 [Herbin 98] 28
 Hessematrix 48
 Hintergrundanpassung 61
 hochauflösende Kamera 106
 Hochdynamikkamera 151, 161
 hof_191096 5, 45, 63
 [Horaud & Sossa 95] 21
 Hough-Transformation 103
 [Hsu & Aquino 99] 33
 [Huang et al. 00] 32
 [Huttenlocher et al. 93] 16, 25
 [Huttenlocher et al. 99] 25
- I**
 IEKF 46, 47
 iterativ erweiterter Kalman-Filter 46, 47
- J**
 [Johnson & Hebert 99] 19
 [Joseph 99] 17
 [Jurie 98a] 17, 18, 31
 [Jurie 98b] 16
 [Jurie 99] 17, 18, 31
- K**
 [Kakadiaris & Metaxas 00] 31
 Kalman-Filter 3, 7, 46, 48, 60, 115, 123,
 143, 155, 159
 Kantenelementbild 6, 7
 Kantenelementextraktion 6
 Kanteninformation 3, 7, 64, 94
 [Kassim et al. 99] 23
 [Kiryati et al. 00] 23
 [Koller 92] 47
 Komplexität des Programms 128
 Komplexität einer Testdatenmenge 40
 Konditionszahlen 48
 konjugierte Gradienten 45
- Kontrast 15, 40, 50, 61, 63, 68, 73, 78,
 81, 83, 90, 93, 124
 Kovarianzen 122, 153
 Kovarianzmatrix 41
 Kreisbewegung 133, 136, 139
 kugelförmiger Behälter 133, 156
 kugelförmiger Behälter im Motorraum
 des Versuchsfahrzeugs 134
- L**
 Latenzzeit 35, 142
 [Lee & Bailey-Van Kuren 00] 34
 [Lee et al. 99] 29
 Leistungsmaß 151
 [Lengyel 98] 19
 [Leuck 00] 145, 151
 Levenberg-Marquardt-Optimierung 47
 [Li et al. 98] 35
 lokale Minima 95
 lokales Minimum 27, 46, 61, 94, 102,
 105, 106, 159
 Lokalisation 11, 93
 Lokalisationsparameter 107
 Lokalisationsrate 151
 Lokalisation, Zusammenfassung 158
 [Lo & Tsai 97] 24
 [Lowe 99] 14
 l_rot 5
 [Lu et al. 00] 17
 Luftfilter 7, 8, 108, 109, 134, 156
 Luftfilter im Motorraum des Versuchs-
 fahrzeuges 134
 Lupe 90, 153
- M**
 [Marchand et al. 99] 33, 35
 Matrixinvertierung 45
 maximaler Anpassungsfehler 38
 MiniVISTA 6
 MIN_PIXELZAHL_KANTE 98
 [Mirmehdi et al. 99] 16
 Modelländerung 74

Modellierung 3, 6
 Modellprojektion 3, 6
 Motorraum des Versuchsfahrzeuges 134
 [Müller & Nagel 99] 127
 multiokulare Lokalisation 104
 [Murase & Nayar 95] 25

N

N 98
 [Nagel et al. 99] 151
 [Nagel & Gehrke 98] 115, 116, 117
 Namen der Bildfolgen 4
 Namensgebung 38, 93
 natürlicher uniformer Spline 138

O

objektzentriertes Objektkoordinatensystem 105
 offene Splinekurve 138
 [Olson 00] 27, 31
 [Olson 98] 27, 31, 94
 [Olson 99] 12, 24, 103
 oszillierendes Verhalten 41
 [Otte 94] 90, 116

P

$\Delta\phi$ 95
 ϕ_{init} 95
 φ_{max} 96
 P 99
 ϕ_{range} 95
 $\Delta\psi$ 95
 ψ_{init} 95
 ψ_{range} 95
 parallele projizierte Modellkanten 115
 Parallelräume im Zustandsraum 156
 Parametrisierung 37, 40, 64
 partielle Verdeckung 124
 [Paulik & Wang 98] 18
 Personenverfolgung 1, 3, 37
 pessimistischer Algorithmus 99
 p_{init} 102, 106

PIXELGENAUIGKEIT 101
 \hat{P}_k^- 43
 [Porill et al. 88] 13
 Portierung 129, 142
 Positionierungsgenauigkeit 101
 Präprozessor 125
 [Press et al. 92] 45, 47
 Programmgröße 128
 Programmkomplexität 128
 Programmname 38, 93
 Prozedurstruktur 144
 Prozeßstruktur 144

R

Θ_R 114
 [Ratan et al. 00] 21
 [Ratan et al. 98] 21
 RAUSCHEN_GRADIENTENRICHTUNG 96
 Realwelt-Anforderungen 93
 [Rechtin 97] 10
 Referenzversuche 4
 Regelung bei der Personenverfolgung 3
 Regelung des Greifers 7, 148
 Regelung des Schraubers 130, 148
 rekursives System 155
 Roboterzelle 4
 Robustheit 2, 9, 34, 35, 37, 38, 40, 58,
 60, 64, 92, 93, 95, 106, 114, 115,
 121, 122, 124, 128, 133, 139, 140,
 154, 160
 Robustheit einer Modellanpassung 40
 Robustheitsschwelle 114
 [Rosin 99] 15, 35
 [Ruf & Horaud 00] 33

S

Θ_S 39
 [Schäfer 96] 125
 Schatten 40, 86
 Schiffsteil 105, 106, 124, 133
 Schrauber 6, 114
 Schrauberregelung 114, 130

Schwingungseffekt 41, 45, 156
 [Sclaroff & Pentland 95] 19
 Sicherheit 35, 92, 124, 128, 155, 160
 [Sidenbladh et al. 00a] 32
 [Sidenbladh et al. 00b] 32
 Sigmoidfunktion 119
 [Silberberg et al. 84] 23
 [Sim et al. 99] 20
 s1_191096 5
 Sockel-Kanäle 150
 Spezialrechner 125
 Spezialrechnerverschaltung 126
 Splinekurve 136
 Stereolokalisation 104
 Suchbereich 96
 Suchbereich einer projizierten Modell-
 kante 96
 Suchschritte 95
 [Sullivan et al. 00] 26
 [Sullivan et al. 99] 27
 Symmetriefehler bei der Kovarianzma-
 trix 44, 46
 Systemkovarianzen 153
 Systemzustand 41
 Szenenbereich 1, 3, 94
 Szenenraumsuche 94

T

$\Delta\theta$ 95
 θ_{init} 95
 θ_{range} 95
 Teilsuche 101
 [Teubner 96] 45
 Θ 38
 Tiefpaßfilterung 90
 Toleranz von Fehlerquellen 105
 Toleranz von Störeinflüssen 93
 [Tonko 97] 4, 6, 7, 32, 92, 115, 122, 123,
 124, 125, 129, 133, 136, 142, 143,
 144, 145, 146, 147
 [Toth & Culverhouse 99] 28
 Trajektorien bei der Verfolgung 156

Transformation der GLST-Eigenwerte
 117
 Transputer 125
 trapezförmiger Suchbereich zu den pro-
 jizierten Modellkanten 96
 Trennung zwischen unwesentlicher und
 struktureller Grauwertvariation
 120
 trinokulare Kameraanordnung 13, 131

U

Übersetzer 90, 125, 129, 143, 145, 151
 Übersetzungsdauer 128
 uniformer Spline 137
 Unschärfefaktor bei der Verfolgung 156
 Unterschiede zwischen den Teilberei-
 chen 2

V

[Vel & Aeberhard 00] 24, 26
 Verdeckungen 40, 49, 61, 73, 78, 124
 Vereinheitlichung der Parameter 64
 Verfeinerung der Suche 99
 Verfolgungsparameter 65, 89
 Verfolgung, Trajektorien 156
 Verfolgung, Unschärfefaktor 156
 Verifikation 13, 14, 16, 17, 19, 31, 95,
 106, 161
 Verifikationsfunktion 95, 106, 161
 Verschaltung der Spezialrechner 125
 Verschiedenartigkeit 40
 Versuchsumgebung Roboterzelle 4
 verteilte Berechnung 35
 verteilte Berechnung der Lokalisation
 148
 [Vijayakumar et al. 98] 18
Visdom 93
 Vollbilder 106, 129

W

[Wachter 97] 3, 4, 5, 32, 37, 41, 43, 44,
 46, 48, 57, 64, 65, 66, 70, 77, 87

[Wachter & Nagel 99] 3, 37, 46
[Wang 99] 29
Warnungen des Übersetzers 90, 125
Wartbarkeit 128, 153
Warteschlange 144
[Weiss & Ray 98] 15
Werkzeugroboter 4
[Wheeler & Ikeuchi 95] 12, 27
[Wiedemann et al. 98] 35, 36
Winkelbereich 96
[Winkeler et al. 99] 25
[Worthington & Hancock 00] 21

X

Δx 95
 x_{init} 95
 x_{range} 95
Xtaro 38

Y

Δy 95
[Yáñez-Suárez & Azimi-Sadjadi 99] 28
[Yang & Cohen 99] 13, 14
[Yang et al. 00] 26
[Yeasin & Chaudhuri 00] 34
 y_{init} 95
 y_{range} 95

Z

Δz 95
Zählung der geeigneten Kantenelemente
102
[Zergeroglu et al. 99a] 33
[Zergeroglu et al. 99b] 33
 z_{init} 95
 z_{range} 95
Zusammenfassung der Lokalisation 158
Zusammenhangskomponenten bei den
Trajektorien 156
Zustandsraum, genutzte Parallelräume
156
Zustandsvektor 156

Zuverlässigkeit 2, 9, 35, 37, 90, 92, 124,
126, 128, 155, 160
Zwangssymmetrisierung 45
Zykluszeit 35, 142
Zykluszeit Bildeinzug 6